

Critical politics of carbon sinks

Last week's suspension of climate-change negotiations in The Hague highlights a political need to grapple with the uncertain science of carbon sequestration before talks resume next year.

So near and yet so far. Seldom can the phrase have been more aptly used than for the failure of a last-minute compromise at last week's round of intergovernmental negotiations on limiting the impact of global warming (see page 503).

What did emerge clearly from the meeting was that the scientific complexities that surrounded the key question addressed at the previous meeting in Kyoto three years ago — whether it is necessary to take steps to limit carbon emissions — pale dramatically compared with the task of negotiating an agreement on the mechanisms needed to put such limits into effect. And, even more so, on the monitoring, evaluation and verification procedures required to ensure that these mechanisms operate effectively and in harmony with other social and economic objectives.

Given such complexities, which some describe as even more labyrinthine than global trade negotiations, it is hardly surprising that the negotiations did not achieve more. Indeed, conference officials' explanation that delegates had "run out of time" has more than the usual ring of truth to it. And the formal description of the meeting as having been "suspended" — hopefully to be continued next spring, probably in Germany — seems a more accurate description than suggestions that the negotiations have broken down irrevocably.

There is one thing that this 'time out' will allow which is essential for progress: a fuller appreciation by all parties, in the context of possible deals to resolve the current impasse, of what we already understand about the role of the terrestrial biosphere in the carbon cycle, and the dynamics of this role as both a source and a sink of atmospheric carbon. There is growing scientific evidence that the contribution of forestry and farming practices to the global carbon budget is of sufficient magnitude to offset some of the carbon released by burning fossil fuels. Such evidence has been eagerly seized on by those countries — particularly the United States — that see in the manipulation of carbon sinks a more attractive way of achieving the overall carbon-emission reductions agreed at Kyoto than cutting down on fossil-fuel use.

Farm factors

It could, for example, lead to the US farming community receiving substantial subsidies to change its land-management practices, such as reducing the depth of ploughing to decrease the release of carbon into the atmosphere. And, given the influence of the 'farm vote', this itself could boost the prospects of an eventual agreement being ratified in the Senate. Indeed, a report earlier this year from the Pew Center on Global Climate Change, *Land Use & Global Climate Change*, points out that, under appropriate conditions, "The Kyoto Protocol can provide incentives for improved management of the terrestrial biosphere" (see: http://www.pewclimate.org/projects/land_use.cfm).

But there is a problem here, namely the many scientific uncertainties that still surround our understanding of the terrestrial carbon cycle. Issues in need of close attention range from suggestions that global warming can itself impair the ability of land and the oceans

to act as carbon sinks, to a lack of detailed knowledge about the long-term effects of short-term measures (such as altering ploughing depths). Indeed, such uncertainties have frequently been used to attack the whole concept of including carbon sinks in post-Kyoto calculations.

This has placed environmentalists in the unusual situation of demanding 'sound science' before action is taken, a demand previously voiced by their opponents, those sceptical of claims of a human impact on global warming itself. In both cases, the demand has had a political motive. But in both cases, too, it reflects the political reality that agreement on action, particularly when substantial costs are attached, is likely to be easier to achieve if it can at least be based on some degree of scientific consensus.

Panel's purpose

Achieving that is a task for a body that was not much in public evidence in the final showdown in The Hague, but which continues to play a key role behind the scenes — the Intergovernmental Panel on Climate Change (IPCC). It was the conclusions of IPCC working groups about the severity of the impending climate change that prompted agreement on the need for action in Kyoto. Since then, further research has strengthened those conclusions. But other working groups have been focusing equally hard on possible mitigation strategies — including the opportunities opened up by what is known as "Land Use, Land-Use Change and Forestry" — or LULUCF (see <http://www.ipcc.ch/pub/srlulucf-e.pdf>).

If last week's meeting had ended in agreement, the next step would have been a request to the IPCC from the ministers present to take on greater responsibility for reporting on such possibilities. The conference's Subsidiary Body for Scientific and Technological Advice had already prepared a proposal that the IPCC should be asked to develop methods for measuring the role of LULUCF activities in the carbon balance, to "prepare a report on good practice guidance and uncertainty management" related to such activities, and even "to examine the feasibility of developing and the implications of applying biome-specific forest definitions".

Even without official blessing such studies are still needed — perhaps more than ever. No one believes that they will by themselves resolve the current impasse — to believe that would be to ignore the wide ideological chasm that separates those in the US negotiating camp, keen to preserve what they can of a market-based approach to climate problems, from their more hard-line European critics demanding tough government intervention. And it remains true that strategies based on managing carbon sinks may be only a short-term measure. In the long run, the only solution is a substantial reduction in overall greenhouse-gas emissions, and clearly one cannot wait for scientific certainty before taking significant steps in that direction. But for now, a key task in the lead up to the continuation of the negotiations next year is to exploit land-use strategies, scientific uncertainties and all, as a means of addressing the political constraints on the United States' ability to mitigate man's impact on climate change. ■





Dante's inferno
Argentina's scientists rebel over plans for agency reform
p504



Planet hopping
NASA under fire for cuts to planetary exploration
p505



What's the beef?
Germany in shock as BSE arrives on its doorstep
p506



Heading south
Britain set to join European Southern Observatory
p507

Deadlock in The Hague, but hopes remain for spring climate deal

David Dickson, The Hague

Supporters of an international agreement to combat global warming are optimistic that more talks next spring could achieve a breakthrough on the role of carbon 'sinks' — the issue that last weekend sank the latest round of climate-change negotiations in The Hague.

Disagreement centred on whether countries' carbon emissions should take account of the contribution of biomass — such as forests and farm crops — in absorbing carbon from the atmosphere.

The meeting had been held to reach agreement on methods of attaining the carbon-emission reductions agreed three years ago in Kyoto, Japan. The Kyoto meeting had led to a consensus that in principle countries would aim to reduce their carbon emission to 5% below their 1990 level by the year 2012, although no major industrialized country has yet ratified this commitment.

Last weekend's breakdown occurred when several European countries rejected a last-minute deal under which the United



Time out: Jan Pronk (above) could not reconcile Jürgen Trittin and Dominique Voynet (right) to the UK-US deal.



States would have agreed to scale back earlier demands to use calculations involving carbon sinks to achieve its Kyoto target.

The breakdown was widely condemned by environmental groups and some scientists, who had been expecting a firm, practical commitment to emerge from the meeting, attended by 7,000 participants from some

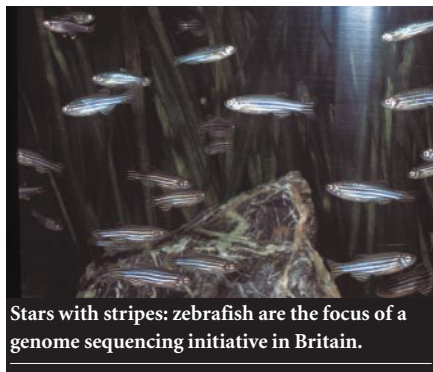
180 countries. Many blamed it on a lack of willingness by political leaders, particularly those in the United States, to take the tough measures needed to reduce global warming.

But several people close to the negotiations say that the gap between the two sides is not large. This was particularly true after the US delegation had backed off from trying to

Wellcome Trust funds bid to unravel zebrafish genome

Declan Butler, Paris

The UK Wellcome Trust is set to fund an effort, costing £5 million (US\$7 million) in the first year, to sequence the genome of the zebrafish (*Danio rerio*) at the Sanger Centre near Cambridge. All the data will be made freely available in the public domain.



Stars with stripes: zebrafish are the focus of a genome sequencing initiative in Britain.

The zebrafish is a model organism for the study of vertebrate development. The fish also has transparent embryos, allowing developmental abnormalities to be spotted by eye. The female lays 200–300 eggs each week — enough for the identification of candidate genes and positional cloning, and the embryos mature within a couple of months. Millions of fish can also be bred in a small space.

The project emerged from discussions last year between John Sulston, then director of the Sanger Centre, and Christiane Nüsslein-Volhard, of the Max Planck Institute for Developmental Biology. Nüsslein-Volhard recently launched a complementary, \$10 million project called the Tübingen 2000 Screen, which should allow point mutagenesis of the entire zebrafish genome to yield embryos with developmental abnormalities. The

US National Institutes of Health may spend US\$5 million on a similar screening project.

Fish genomes have long regions of 'synteny', where the order of the genes is similar to that found in humans, says Leonard Zon, a zebrafish researcher at the Children's Hospital of Boston. "This will allow genome 'ping-ponging,'" he says. "If a zebrafish mutant falls in a region of synteny, the [corresponding] local region of the human genome can be searched."

The zebrafish genome has about 1.7 billion base-pairs. The full sequence and the massive mutant screen should allow every developmental gene to be isolated, from which the mutated genes can then be cloned. Many of the genes isolated will be relevant to human systems, so that research on zebrafish mutants will assist the annotation of the human genome.



US delegate Frank Loy agreed to compromise.

This move was reported by the *Washington Post* as having followed a late-night conversation between President Bill Clinton and British Prime Minister Tony Blair, and formed the basis of a compromise agreement thrashed out between Blair's deputy, John Prescott, and the US delegation led by Frank Loy.

But some European countries rejected the agreement. Particularly strong opponents were the environment ministers of France and Germany, Dominique Voynet and Jürgen Trittin, respectively, who are both members of their countries' Green parties. Voynet said that Britain had "conceded too much to America".

Supporters of the US-UK compromise blamed the failure to bridge the negotiating gap on the inability of government officials, exhausted after two weeks of talks and two all-night negotiating sessions, to calculate its technical implications accurately.

Jan Pronk, the Dutch environment minister and chairman of the conference, said that the key political issues — including rules for counting emissions from carbon sinks — "could not be resolved in the time available".

Environmental groups oppose any allowance for sinks in emissions calculations. Greenpeace described earlier proposals from Pronk, which would have allowed the United States to include 50 million tonnes of sinks a year in its calculations, as a "free gift" that would allow emissions to continue growing.

But there are signs that the inclusion of sinks could increase the chances of the Kyoto agreement being ratified in the US Senate. US farmers' groups this month came out in favour of an international agreement on carbon sequestration.

"This is a critical shift which greatly improves the chances that the Senate will ratify the Kyoto Protocol," says Philip Clapp, president of the National Environmental Trust, a Washington-based advocacy group.

Margot Wallström, the European Union's environment commissioner, said that the experience of the UN Convention on Biological Diversity, which was agreed only at the second attempt, "shows that a temporary setback can lead to a better result in the end". ■

include reforestation efforts in the developing world in its calculations, and had cut its demands for the maximum domestic allowance for carbon sinks from 300 million to 75 million tonnes a year.

Argentinian researchers fight government plans for reform

Xavier Bosch

Argentinian scientists, already facing financial constraints caused by the country's international debt crisis, are rebelling over the government's plan to reform the National Council for Science and Technology (CONICET).

One of Latin America's most prestigious research agencies, CONICET employs 3,500 of Argentina's best researchers. Earlier this year, the Argentinian cabinet approved a plan by the science and technology ministry to boost links between CONICET and the country's universities.

But senior Argentinian scientists have attacked the plan as financially motivated and detrimental to CONICET's hard-won scientific reputation.

Under the plan, all new members of CONICET would join at the same rank and would receive a \$1,000 supplement to any university salary. They would compete with each other for university promotions, but if any of the new researchers failed to make the grade after four years, they would lose their CONICET membership.

Evaluation of the researchers' work would be carried out jointly by the universities, CONICET and the National Agency for the Promotion of Science and Technology. The status of current CONICET members would be unchanged by the reforms, although they could opt to join the new system.

The reforms will give more opportunity to young researchers and increase their mobility, says Dante Caputo, secretary of science and technology. He believes that CONICET has become too bureaucratic, and that

careers are blocked by its hierarchical structure, which divides researchers into five ranks based on seniority.

In a statement, the directors of some of the agency's research centres admitted that "young people have almost no possibility of entering CONICET". The average age of a CONICET researcher is 50, a problem the directors blamed on low public investment.

Armando Parodi, scientific director of the Biotechnology Research Institute at San Martin University near Buenos Aires, says that CONICET promotions are evaluated seriously and based on academic merit. He says that Caputo's plan would destroy the agency's career structure, and that it was drawn up without consulting researchers.

Parodi adds that the plan is supported by the administrators of the country's big universities, who he believes want to get their hands on CONICET's funds and its influence on academics. He says that the government's argument for better interaction does not stand up, as three-quarters of CONICET researchers already teach in universities.



Young scientists will benefit, says Dante Caputo.

José Viramonte, director of CONICET's Geonorte Earth sciences institute in Buenos Aires, says that the plan has been devised by Argentina's "bureaucratic establishment" to extend its power.

"Unlike the universities, CONICET has kept its distance from electoral commitments and cronyism," he says. ■

German universities used forced labour

Christina Hohmann, Munich

Academic institutions in Germany used thousands of forced labourers during the Second World War, according to evidence emerging from some universities and the Max Planck Society.

Victims of the practice, and in some cases their relatives, are eligible for compensation from a DM10 billion (US\$4.3 billion) fund created last year by German industry and government. But time is hardly on their side: only about 2.4 million of the estimated 11.3 million forced labourers who worked in Germany between 1939 and 1945 are thought to be still alive.

Discussions in the press about large-scale forced labour in industry have tended

to overshadow the actions of universities, which were also major employers at the time. But some universities have searched their archives for evidence of culpability.

The University of Tübingen, for example, has admitted that it used 150 forced labourers, 120 of whom were eastern European women working in kitchens or as nurses in the university clinic. The men, mostly prisoners of war, worked as handymen or gardeners.

The University of Heidelberg's records from 1944 list 23 'eastern workers', as well as 17 Russian civilians, seven French workers and one Croatian. The University of Stuttgart's list of alien workers and prisoners of war, drawn up in 1946 at the

NASA U-turns enrage planetary scientists

William Triplett, Washington

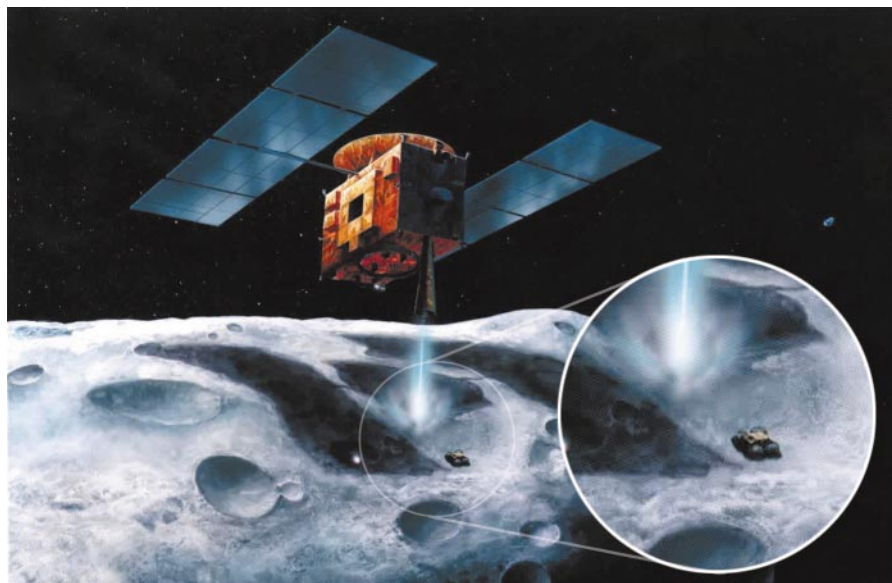
Planetary scientists in the United States have criticized a string of recent cancellations of missions at NASA, warning that they may put the future health of US planetary exploration at risk.

Following NASA's fourth cancellation of a scheduled planetary science mission in two years, the American Astronomical Society's Division for Planetary Science (DPS) has issued a strongly worded statement arguing that any more cancellations would result in "excessively reduced" programme content and narrow the scope of US space research.

The latest cancellation was the nanorover, a NASA component of Japan's MUSES-C asteroid sample-return mission. That followed the abandonment of the DS-4/Champollion comet mission and the Mars 2001 Lander, a follow-up to the Mars Polar Lander that failed in late 1999. In September, NASA stopped work on the Pluto-Kuiper Express mission, scheduled to launch in 2004, that was to study Pluto's atmosphere before exploring the Kuiper Belt.

NASA says the cancelled missions were plagued by cost overruns. But the planetary scientists say NASA could have tackled these by increasing the peer review of proposed missions and making the mission-proposal process an open competition. For decades, the Jet Propulsion Laboratory (JPL) in Pasadena, California, has been in charge of NASA's planetary exploration missions.

Ed Weiler, NASA's head of space science since 1998, who cancelled three of the four missions, says he would like to implement more competition. But Weiler maintains that he must first cope with the problems facing existing missions, which were all instigated by his predecessor, Wes Huntress, current vice-chairman of the DPS.



Rover's over: the nanorover (inset) on the MUSES-C asteroid mission is NASA's latest cancellation.

Weiler argues that the cancellations were unavoidable. The nanorover, for example, was priced at around \$20 million about four years ago, but the latest estimates put its cost at more than \$60 million. Similarly, the JPL put a figure of \$650 million on a two-mission package including the Pluto expedition, whereas the latest figures put the price at over \$1.4 billion.

Michael Drake, who chairs the solar system exploration subcommittee of NASA's Space Science Advisory Committee, says he is "sympathetic to NASA on this because fiscal discipline must be maintained". The space agency has been plagued by unrealistically low bids from contractors, says Drake.

But Mark Sykes, chairman of the DPS, says Weiler's hardline approach risks "throwing the baby out with the bathwater". He fears that the Europa orbiter — a mission to

explore one of Jupiter's moons for signs of life — could be next for the chop.

Sykes acknowledges that NASA's plans for exploring Mars are still very much alive, but he does not want Mars to be the only planet studied.

Weiler concurs, but says: "I want to find a way to make these other outer-planet missions cheaper." He says he is considering opening missions up to bids from universities or federally funded institutions. Weiler has told the JPL that the Pluto mission could go ahead if it can cut costs on other projects.

Huntress and his allies would like to see NASA consider alternatives to cancelling missions. "There are lots of ideas out there," he says. "If one mission is running over costs, why not look for other cheaper ideas for the same mission?" ■

request of the local government after the war, contains 56 names.

These lists are almost certainly



Captive workforce: forced labourers in the University of Stuttgart's library in 1943.

underestimates of the true number of forced labourers. Dieter Speck, head of the archive of the University of Freiburg, which lists more than a dozen cases there, says that investigations are "very difficult and frustrating", as many records have been destroyed or handed over to the occupying forces. Extrapolating from surviving documents, he estimates that academia as a whole must have exploited several thousand forced labourers.

The Kaiser Wilhelm Society, the predecessor of the Max Planck Society, employed a "large number of forced labourers, maybe up to a thousand individuals", says Jens-Christian Wagner, a historian and member of an independent

research group investigating the episode. Its findings will be published next month.

Researchers have found that a class system operated among the forced labourers. Those from the west were often considered capable of being 'Germanized' and received preferential treatment. For example, three French prisoners of war worked at the University of Freiburg in their own professions as librarians and archivists, and remained at the university after the war.

Most universities have taken few steps to contact forced labourers or their relatives. One exception is the University of Tübingen, which some years ago invited 30 former forced labourers from Poland to a reception as a gesture of apology. ■

Dispute over rock dating settled out of court

Rex Dalton, San Diego

An Earth scientist from Columbia University in New York has settled a lawsuit filed against him last year by a geographer in Tempe, Arizona.

The dispute between Wallace Broecker, a professor of Earth science at Columbia's Lamont-Doherty Earth Observatory, and Ronald Dorn of Arizona State University (ASU) began when Broecker co-authored an article in *Science* questioning Dorn's rock-dating techniques.

Three weeks ago, Broecker and his university made a confidential settlement with Dorn. The amount of the settlement was not disclosed, but sources say it is between \$30,000 and \$70,000.

Declining to discuss the case, Broecker issued a brief statement: "[We] have resolved the litigation and have agreed that any scientific disagreements ... should be debated in the scientific arena and not addressed through litigation." Neither Dorn nor his Phoenix-based attorney would comment.

In June 1999, Dorn sued Broecker and his co-authors from the University of Arizona in Tucson, Northern Arizona University (NAU) in Flagstaff and a Swiss institute for defamation, intentionally causing emotional distress and intentional interference with his work.

Dorn's lawsuit against the remaining scientists is continuing in state court in Phoenix. The defendants are expected to seek its dismissal.

The lawsuit was based on an article that claimed that Dorn's samples of rock varnish appeared to contain bits of coal and charcoal, which when combined could distort radio-carbon tests (*Science* 280, 2132–2139; 1998). Dorn's technique involved scraping varnish off rocks, treating it and then submitting the samples for radio-carbon dating.

A complaint about Dorn's rock-dating procedures was also made to the National Science Foundation (NSF), which had funded some of his work. The NSF investigation was closed after an ASU inquiry cleared Dorn in September 1999 of fabricating data in published articles (see *Nature* 401, 629; 1999).

Dorn continues to work at ASU, where he is part of a team of 20 scientists seeking funding for information-technology studies associated with ASU's NSF-funded Long-Term Ecological Research Project. ■

Germany rues 'complacency' over BSE testing strategy

Alison Abbott and Quirin Schiermeier, Munich

The German government has been prompted to order extensive tests on cows following the identification last week of two cases of BSE (bovine spongiform encephalopathy) in German-born cattle. Within the next few weeks it should learn whether these cases represent only the tip of an iceberg.

Many scientists fear that the true extent of the disease may have been hidden. Germany had declared itself 'BSE-free', largely because its farmers have not traditionally fed their cattle on the blood and bone-meal that are thought to have sparked the crisis in Britain.

The precautions taken have therefore not been as rigorous as in other countries. When Britain banned the use of bone-meal as feed in all farm animals, for example, Germany banned its use only in cattle, allowing it to continue in pigs and poultry.



Moreover, Germany has only carried out BSE tests on animals showing symptoms of central nervous system disturbance — there have been around 15,000 such tests in the past ten years.

The federal government has now called for testing of all cattle at high risk of developing BSE, including those dying of unknown causes. Such cases average 66,000 per year. Cattle slaughtered over the age of 30 months might also be added to the list, because older cattle have a higher risk of accumulating the infectious prions thought to cause the disease.

In the wake of last week's discovery, some scientists say that German complacency was misplaced. Hans Kretzschmar, a prion expert at the University of Munich, and a member of an ad hoc group of experts called on occasionally to advise the government, regrets that scientific advice on the matter was not institutionalized in Germany. "There was a feeling that Germany could never be affected by BSE, so a standing advisory committee would not be necessary," he says.

"But when an animal born in Germany in 1996 contracts BSE, even though the use of blood and bone-meal was banned in cattle in 1994, you start to wonder what can be believed," Kretzschmar says. He warns that variant Creutzfeldt–Jakob disease, the human form of BSE, may now arrive in Germany.

"We have been too complacent in assuming that Germany was immune," says Reinhard Kurth, head of the Robert Koch Institute, the government agency that researches into infectious diseases. Kurth regrets what he sees as the general lack of a sound scientific basis in dealing with BSE issues. ■

Chinese science goes to Earth

David Cyranoski

Earth sciences will be the focus of a concerted push by the Chinese Academy of Sciences, which last week formally opened its new Institute of Geographic and Natural Resources Research.

Formed by the merger of two older institutes, the new institute is part of a wider academy reorganization announced two years ago (*Nature* 394, 710; 1998).

About 400 scientists will join the institute from the old Institute of Geography. Of those, about 250 will take part in a new incentive programme, designed to encourage a more competitive approach to research.

Scientists can apply to be part of this programme "if their research interests align with

the institute's strategy", according to Li Xiubin, the institute's deputy director. He adds that this will entail a strong emphasis on basic research. Projects inside the incentive programme will be evaluated every two years to determine whether they can continue.

Other employees of the former institutes will be moved to a company providing services to the new institute, which will eventually be spun-off from it.

The academy's old organization was largely modelled on that of the former Soviet Union during the 1950s and was badly in need of reform. Now it plans to streamline its 123 institutes — which house many of China's best scientists — and reduce its staff from 68,000 to 30,000 by 2010. ■

Britain plumps for support of big projects...

Peter Aldhous, London

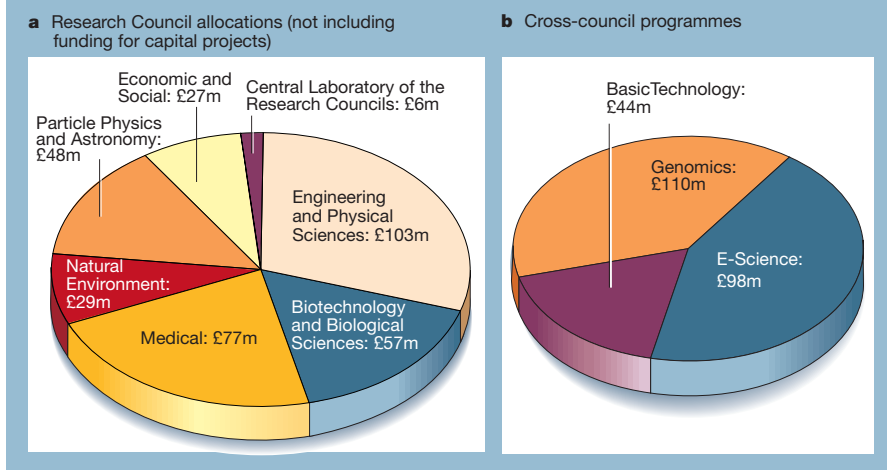
Researchers working in priority areas such as post-genomics and information technology had cause to celebrate last week as the British government published the fine print of its comprehensive spending plan for science.

But despite overall increases in funding of about 7% for each of the next three years, including spending on buildings and equipment, those working in less fashionable disciplines could be left out in the cold.

"The probability that your average project grant will be funded is not going to go up," warns Peter Cotgreave, director of the lobby group Save British Science. Indeed, last week's announcement is the most obvious development in an emerging UK trend: the concentration of resources in big, collaborative programmes, rather than increasing the funding for individual grants.

"An inevitable consequence of the way these programmes will work is that more money will be concentrated into a smaller number of groups," says David Clark, science and engineering director at the Engineering and Physical Sciences Research Council.

The new money emerged from a strategic review of planned government spending for the three years beginning 2001. In July, the government announced the size of the overall increase to the science budget, plus an additional fund of £1 billion (US\$1.4 billion) to update laboratory equipment and buildings — £225 million of which will be pro-



In the money: how the new funding will be allocated across disciplines over the next three years.

vided by the Wellcome Trust (see *Nature* 406, 335; 2000). But it was not clear how the increases for research would be spent.

This extra money totals £356 million over the three years. But £252 million is earmarked for cross-research-council projects in genomics, 'E-science' and what the government calls 'basic technology' (see figure). The rest is divided between the research councils roughly in proportion to their current spending. After each council has funded its own priority projects, there will be little left: indeed, some councils will have to cut back on their existing activities (see below).

The main thrust of the genomics initiative will be to determine the function of genes that have been sequenced. George Radda, chief executive of the Medical Research Council, expects the focus to be on structural genomics and population-based studies. He says the comprehensive medical records gathered by the National Health Service make Britain an ideal place to study the links between genes, lifestyle and disease. The Medical Research Council plans to spend £20 million on the UK Population Biomedical Collection, a national database involving some 500,000 volunteers, in collaboration with the Department of Health and the Wellcome Trust.

The E-science initiative will focus on developing a distributed computing concept known as the Grid, and applying it to climate modelling, bioinformatics and handling the expected deluge of data from the Large Hadron Collider at CERN, the European laboratory for particle physics in Geneva.

The basic technology component includes research in nanotechnology and robotics. Officials at the engineering research council expect particular emphasis to be given to photonics — computing and data transmission using photons instead of electrons — where Britain is an acknowledged world leader.

The architects of the government's science policy make no apologies for these earmarks. "We definitely do need things that are larger and more collaborative," says Robert May, who was the government's chief science adviser until October, and this week takes over as president of the Royal Society.

Most researchers agree that such work is important, but some fear that the trend to attach strings to science spending may grow at the expense of basic, curiosity-driven research. "I wouldn't yet say this is a big problem, but there is a potential for a problem," says Cotgreave. "We've got our eye on it."

http://www.dti.gov.uk/ost/whatsnew/sciencebudget.pdf

...but ESO membership comes at a price

Britain will join the European Southern Observatory (ESO) in 2002. Last week's science spending plan includes funds for the Particle Physics and Astronomy Research Council (PPARC) to pay for ESO membership.

The move will give British astronomers access to the new Very Large Telescope, which has four 8.2-metre mirrors, at Mount Paranal in Chile. This will require a 'joining fee' of £70 million (US\$98.5 million), plus £12 million per year.

But the timing of the payments must be decided before British astronomers learn what other projects must be sacrificed so that they can use the ESO's facilities.

The PPARC's budget, which will rise to £232 million by 2003–04, includes £10 million over each of the next three years for membership of the ESO. If the



Joining ESO will give Britain access to Chile's Very Large Telescope.

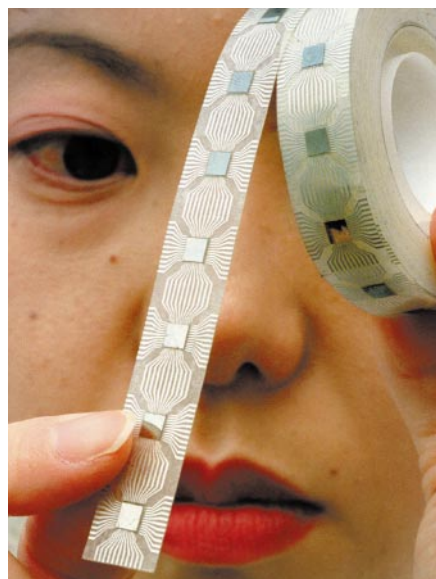
joining fee were to be paid in one go, the consequences for the rest of the PPARC's programme would be disastrous. But in practice a deal will be thrashed out to spread the payments over several years. "There is flexibility on the timescale," says ESO director-general Catherine Cesarsky.

But some cuts to the PPARC's existing research portfolio are

inevitable. "What I hope is that the PPARC will consider its whole programme," says Martin Rees of the University of Cambridge, Britain's Astronomer Royal.

Ground-based astronomy is expected to bear the brunt of any cuts, however, and an expert panel of astronomers has been called in to prioritize the current range of projects.

P.A.



Small in Japan: semiconductor research is on a roll.

Japan plans joint project on nanoscale semiconductors

Tokyo Japanese semiconductor companies have announced plans to join a government-led nanoelectronics research and development effort that aims to build semiconductor devices in the 50-nanometre range by 2007.

Under the initiative from the Ministry of International Trade and Industry (MITI), engineers from large semiconductor companies will join forces with researchers from national laboratories and universities to develop process technologies for the semiconductors.

But overall funding for the project remains unclear, with the government's ambitious spending plan for information technology already meeting some public criticism. MITI officials say they have requested ¥6 billion (US\$54 million) in funding for the project next year, but admit that only a fraction of that amount may eventually be granted.

In parallel to the MITI initiative, the Japan Electronics and Information Technology Industries Association has announced a \$750 million joint project to develop production technologies for 0.1-micrometre processes and system-on-chip technologies — objectives that are almost within reach.

DNA archive in fight against Japanese whale meat sales

Tokyo The Japanese Fishery Agency has announced plans to clamp down on the sale of illegal whale meat by archiving the DNA of captured whales.

In Japan, only a few kinds of whale meat are allowed to be sold, including Minke

whales caught in a controversial research programme, and there is widespread doubt about the source of meat in shops.

The agency plan involves registering DNA from whale skin samples as soon as a whale is captured. Eventually, a database would allow the meat in supermarkets to be traced. Regulating the distribution of whale meat is seen as a step towards overcoming international resistance to the resumption of commercial whaling.

Australians hunt for Tonga's disease genes

London An Australian biotechnology company has secured exclusive access to a possible genetic goldmine on the Pacific island of Tonga. In an agreement brokered by the International Diabetes Institute, Autogen has won the right to search for links to disease in the gene pool of the island's 108,000 residents.

Autogen's main interest is in finding links to diabetes and obesity, both of which are common on Tonga. The company hopes that the Polynesian island's long isolation, along with its good kinship records and preference for large families, will simplify the search for disease-related genes.

The firm has pledged to share the profits from any drugs or therapies that come out of the research with the people of Tonga.

US researcher banned for fabricating data

Washington The US government has banned Evan Dreyer, a University of Pennsylvania researcher, from receiving federal research funds for ten years because he fabricated data in a 1996 grant application while at Harvard Medical School.

The health department's Office of Research Integrity has investigated around 100 misconduct cases, but Dreyer — an associate professor of ophthalmology and director of the university's glaucoma programme — is only the second to receive such a severe punishment.

A spokeswoman for the University of Pennsylvania says the university was unaware of the allegations against Dreyer when he resigned from Harvard and moved there in 1997.

British public doubts genetics rules

London Nearly three-quarters of British people think that government regulation has not kept up with human genetics research, according to a survey released last Monday by the UK Human Genetics Commission.

Almost a third of those surveyed by MORI (Market & Opinion Research International) believe that "genetic research

is tampering with nature and is therefore unethical". But 90% of the respondents expect that human genetics research will result in cures for many diseases, and 70% believe that it will lead to healthier babies.

Support for taking DNA samples from suspects of violent crime was almost unanimous, but less than half of those interviewed were comfortable with applying this in cases of drink-driving or fraud. Only 8% thought genetic test results should be used for setting insurance premiums.

Redundancies looming at John Innes Centre

London Scientists at the John Innes Centre in Norwich could face redundancy next month, when its management reviews the centre's activities in an attempt to free up funding for research into new areas of science.

Just over half of the research centre's funding is currently allocated to specific projects. The remainder — a £10 million (US\$14 million) grant from the Biotechnology and Biological Sciences Research Council — can be used to seed new scientific opportunities. But the centre, which focuses on plant and microbial research, has almost quadrupled in size over the past 10 years, and much of this grant is consumed by operational costs.

Chris Lamb, the centre's director, says he needs to free up £2 million in the next two years. Areas of science less likely to contribute to the centre's growth will be identified, and changes, including redundancies, will be made. The exercise is expected to be complete by mid-December with the loss of about 30 jobs.

European lab sees budget rising

Heidelberg The 16 member states that run the European Molecular Biology Laboratory (EMBL) have approved a significant increase in the laboratory's budget.

Subject to a formal ratification by Denmark, the EMBL council has agreed to increase the lab's budget by 25% over the next five years — a boost of 11.5 million euros (US\$10 million).

The European Bioinformatics Institute, EMBL's outstation in Hinxton, near Cambridge, will get nearly half of the top-up. Its 6.1 million euro baseline budget will increase by 5 million euros. The institute has been underfunded, its supporters say, with the European Commission and European Union member states both avoiding responsibility (*Nature* 405, 723; 2000).

The council also agreed to allow EMBL to establish an endowment fund to support technology transfer and small start-up companies. The fund will be financed from independent private sources.

To the planets on a shoestring

Small space probes can thumb a ride into space and then hurl themselves around the Solar System using orbital gymnastics — meaning that you no longer need bottomless pockets to do planetary science. Robert Adler talks to the thrifty engineers who are making it happen.

Faster, better, cheaper' was NASA's catchphrase for most of the 1990s. But cheaper has come to seem like a false economy. After the embarrassing loss of two successive Mars missions, NASA has redrawn its plans to explore the planet, and is boosting its spending on the programme by about a third. NASA administrator Daniel Goldin, who championed 'faster, better, cheaper', seems to accept that his goal of reducing mission costs led to corners being cut.

Yet some space engineers remain adamant that they could send a flotilla of probes to Mars or other planetary destinations for the cost of just one of NASA's planned missions: their motto could be 'smaller, smarter, cheaper-still'. Using small spacecraft that would hitch a ride into space with commercial satellites, these engineers would save hundreds of millions of dollars in launch costs. Then, by cleverly exploiting orbital trajectories, the probes could reach their targets on a fraction of the normal fuel requirement.

If these launch and trajectory strategies could be combined with miniaturized instruments, 'intelligent' navigation systems and advanced propulsion technologies, they might transform space science from an expensive minority pursuit into an affordable, mainstream scientific activity. "There are a huge number of things to study out there," says Charles Garner, a senior engineer at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California. "But if you're going to study

them to any substantial degree, you've got to have cheap access to space."

Getting into orbit around the Earth can account for up to half of the cost of a planetary science mission. And as long as this is done using chemical rockets, the basic economics are rigid. "Launch costs have not changed significantly in the past 20 years," says Maggie Jones, manager of the small satellite programme of Ball Aerospace, a company in Boulder, Colorado. "So something else has to change."

Given that the cost of getting into space increases with a craft's mass, the most obvious thing to change is a probe's size. It took an excessive amount of cash to get some of the huge spacecraft designed by NASA in the 1980s off the launch pad. Galileo, for instance, now completing its mission to Jupiter and its moons, tipped the scales at 3,881 kilograms; its 1989 launch, on the space shuttle Atlantis, cost \$550 million.

Hitchhiker's guide

When he took over at NASA in 1992, Goldin recognized the problem. Such bloated spacecraft put too many eggs in one basket — as the loss of the \$1 billion Mars Observer probe in 1993 demonstrated. Subsequent Mars missions were slimmed down: Mars Global Surveyor, for instance, now orbiting the Red Planet, had a mass of some 1,000 kg at launch. But to the 'smaller, smarter, cheaper-still' brigade — who are working on spacecraft a fraction of this size — Surveyor and its kind are still heavyweights.

With smaller spacecraft, costs can be slashed by hitching a ride into space on a commercial satellite launch. Several start-up companies hope to exploit this market. For example, EUROCKOT, based in Bremen, Germany, is using modified Soviet SS-19 ballistic missiles, and plans its first launch in late

Hop on: Ariane 5 has room for up to eight small hitchhikers.



2001. The rockets' main job will be to launch commercial satellites of around 350 kg. But the company also offers a 'LaunchaPiggy' service, with piggyback slots for smaller craft costing as little as \$10,000 per kilogram. If each of the piggies weighs 50 kg, seven could go into orbit at a time.

A few of the big players in launch services already offer piggyback rides. Europe's Arianespace consortium has launched 24 small satellites, weighing up to 50 kg each, on its Ariane 4 rockets. The larger Ariane 5, which began commercial operations in December 1999, can carry up to eight 120-kg piggyback payloads per launch. Alternatively, two of the piggyback units can be combined to carry a single 240-kg payload. Because the operator of the main payload — which is usually a large communications satellite — covers the launch costs, the hitchhikers can fly for just one or two million dollars.

Satellites with altitude

So far, piggyback launches have only put small satellites into Earth orbit. But Ariane 5 has started to fire the imaginations of engineers interested in exploring the planets, as a piggyback slot of 240-kg could ferry a respectable package of scientific instruments to Mars or beyond. "Ariane 5 is very exciting for planetary missions," says Jacques Blamont, adviser to the director-general of CNES, the French space agency.

Ariane 5 leaves its payloads in a geosynchronous transfer orbit (GTO). These are highly elliptical, typically coming within a few hundred kilometres of the Earth's surface at their lowest point, or perigee, but extending out to some 36,000 km from Earth at their highest point, or apogee. At this altitude, objects in a circular orbit remain over the same point on the Earth's surface at all times — hence the term geosynchronous. Telecoms satellites move from an elliptical GTO to geosynchronous orbit by using thrusters to boost their altitude at perigee.

Blasting straight from a GTO to the planets would take huge amounts of fuel. But it is possible to leave a GTO much more efficiently and head off in almost any direction. The secret is to use a trick thought up in the late 1980s by Edward Belbruno, an applied mathematician who specializes in celestial dynamics, now at Princeton University in New Jersey, and James Miller, an expert in spacecraft navigation at the JPL.

Belbruno and Miller realized that modest amounts of rocket power could increase the apogee of a spacecraft's orbit until it lies several hundreds of thousands of kilometres from the Earth. At such distances, explains Belbruno, probes can enter a region of space known as the Weak Stability Boundary, where the gravitational pulls of the Earth, Moon and Sun almost cancel one another out. In this zone, the mathematics of chaos rules, and a gentle rocket burn can radically

alter a spacecraft's trajectory. "It's like a drunk walking down the street," says Belbruno. "At any given moment, he can veer off in any direction."

The calculations needed to achieve a particular result are fiendishly complicated. But two space salvage operations have shown that these spacefaring drunks can be steered successfully to their destination. The first involved a Japanese spacecraft called Hiten, launched in 1990, which was supposed to eject a basketball-sized probe called Hagoromo into lunar orbit. Unfortunately, Hagoromo was lost, and mission controllers turned to Belbruno and Miller's calculations to rescue the mission.

Hiten, parked in an elliptical Earth orbit, did not have enough fuel to get to the Moon by conventional means. But by gradually extending the satellite's orbit until it reached more than a million kilometres beyond the Moon — some 1.4 million kilometres from the Earth — a very small thrust was enough to ease Hiten into lunar orbit in early 1992. "Japan became the third country in history to send a craft to the Moon," says Belbruno, "and the first to demonstrate a Weak Stability Boundary transfer."

The second salvage mission resurrected Asiasat 3, a communications satellite that ended up in a useless low orbit after the fourth stage of its launcher failed on Christmas Day 1997. There it stayed until, in May 1998, engineers with Hughes Electronics in El Segundo, California, boosted its apogee to take it just beyond the Moon. After two lunar swingbys, the satellite's thrusters fired, slowing it and altering its trajectory as it swung back towards our planet. Earth's gravity then pulled it into a useful orbit. Renamed HGS-1, the satellite is now in commercial use.

Encouraged by these successes, the European Space Agency (ESA) is planning to use a trajectory based on Belbruno and Miller's concepts to move a spacecraft called SMART-1 from Earth to lunar orbit. Sched-



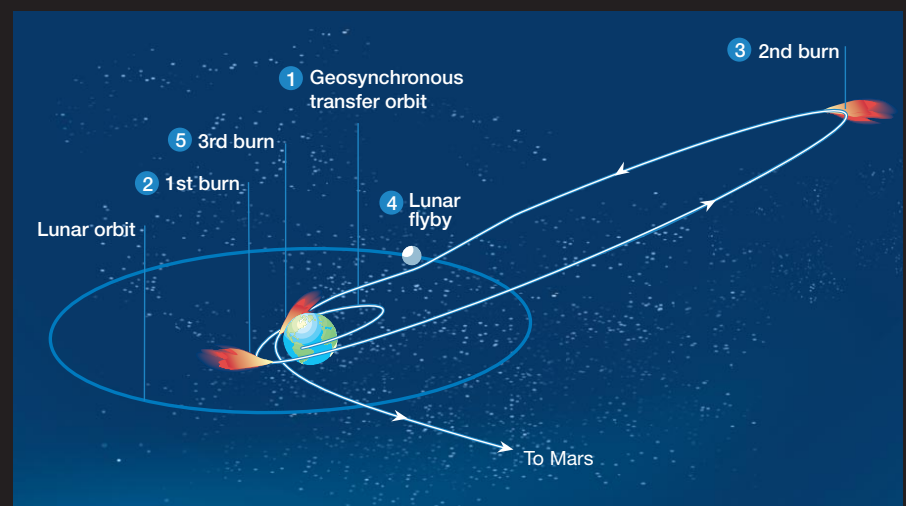
Back on course: Asiasat 3 swung by the Moon before returning to Earth orbit.

HGS

uled for launch by Ariane 5 in October 2002, this craft will take pictures of the Moon's surface while studying its chemical composition using infrared and X-ray spectrometers. SMART-1 is also intended as a testbed for propulsion and navigation technology for future ESA craft including the Bepi Colombo mission to Mercury, scheduled for 2009.

Mars and beyond

But Blamont wants to use the Weak Stability Boundary to get spacecraft to Mars. Working with experts in celestial dynamics, including Paul Penzo of the JPL, by the late 1990s Blamont had convinced NASA that one of Ariane 5's 240-kg piggyback berths could allow 50 kg of hardware to reach Mars — enough to yield good scientific returns, given advances in instrument miniaturization. The propulsion system and fuel would account for most of the remaining mass. Penzo devised an elegant strategy to get from GTO to Mars using just three rocket burns and the gravitational pulls of the Earth and Moon (see diagram, below).



Three burns and you're off: Penzo's fuel-efficient scheme to send a space probe on its way to Mars.

“At first it didn’t look feasible,” says Penzo. “But after about a year I latched onto a good way to do it.” The first burn, at the spacecraft’s closest approach to Earth, stretches the orbit beyond the Moon and into the Weak Stability Boundary. A second burn nudges the craft back towards the Moon, whose gravity adjusts its trajectory once more. Finally, the combination of a third burn, again at closest approach to Earth, and the planet’s gravitational pull, slings the craft towards Mars.

Trips to Mars must begin at just the right time, if the planet’s orbit is not to take it out of a spacecraft’s reach. But although piggyback launches must fit in with the timing desired by operators of the primary payload, this is not a problem: the craft can be parked in a GTO until Mars is in the right place. This approach could deliver a series of missions to Mars for as little as \$50 million each. But in the soul-searching following the losses of Mars Climate Orbiter and Mars Polar Lander in 1999, NASA put the project on hold.

Penzo has also devised a similar way to get to Venus using the Moon’s and Earth’s gravity. This opens up the possibility of sending piggyback spacecraft to Venus, or using the planet’s gravity to swing a probe towards another destination. Despite NASA’s cold feet, Kim Leschly, formerly Penzo’s project manager at the JPL, believes that the idea of using the Weak Stability Boundary to send small spacecraft to the planets is so clever and cheap that it will be resurrected within the next three years. “I’m certain this concept will revive itself,” he says.

The chances of that happening could be enhanced by work on the ‘smaller’ and ‘smarter’ aspects of the strategy. Among the key players are engineers at the JPL’s Center for Integrated Space Microsystems (CISM), which was established in 1998. Its director, Leon Alkalai, explains that the centre’s goal is to miniaturize systems so that they can be

put onto a series of chips that will allow probes to function autonomously.

The CISM is also applying the techniques used to etch silicon chips to make tiny machines. The centre has already made a gyroscope, vital for controlling a spacecraft’s position, that is smaller than a dime. And it is working on devices including star trackers, accelerometers and navigation systems, plus power, command and control, communications, and data-handling units. “These will all be miniaturized to chip-level subsystems, whereas in the past they would have been boxes,” says Alkalai. “We’re talking about at least an order of magnitude smaller.”

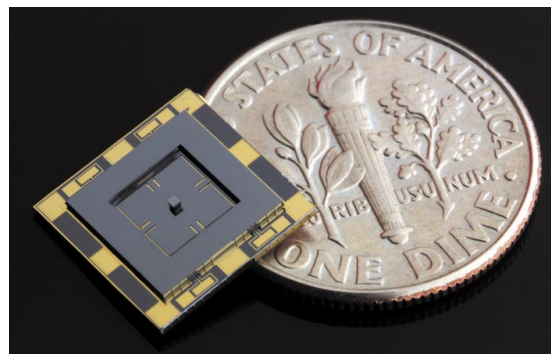
A spacecraft on a chip

Alkalai hopes to fly technology demonstration experiments by 2006. By 2012, he expects to see spacecraft powered and managed by CISM-designed chips exploring the moons of Jupiter or Saturn, or fetching material from a comet back to Earth. And in the longer term, he plans to explore revolutionary computing strategies to create “a thinking spacecraft”, able to integrate the information from multiple sensors, recognize patterns, adapt to rapidly changing environments, and deal with faults or external emergencies.

While the CISM looks to the future, several companies are using today’s state-of-the-art electronics to build low-maintenance craft that need little input from the ground. “People come here and they’re horrified because there’s nobody in the control room,” says Martin Sweeting, chief executive of Surrey Satellite Technology, a British firm affiliated with the University of Surrey in Guildford. The company has put 19 small satellites into Earth orbit since 1981, and is now plotting to send a 420-kg probe to the Moon in 2003. “There’s a large science community interested,” Sweeting says.

Given that rockets and their fuel can account for more than half of a planetary science probe’s mass, finding more efficient propulsion systems is another priority. Many ideas have been proposed. But so far, only one has made the trip from drawing board into space. Solar electric propulsion uses power from a spacecraft’s solar panels to ionize a gas, which can then be expelled from a nozzle at high velocity using an electromagnetic field.

Solar electric propulsion made its debut on NASA’s Deep Space 1 craft, launched in October 1998 and headed for a September 2001 rendezvous with Comet Borrelly. Deep Space 1 is powered by an engine that expels xenon ions at 35,000 metres per second. This high velocity makes the ion engine up to ten times more efficient than a chemical rocket — meaning less fuel and a lighter spacecraft. And although current ion engines generate only a tiny fraction of the thrust of conventional rockets, they compensate by being able to fire for thousands of hours. Encour-



Turn on a dime: NASA’s miniature gyroscope.

aged by the success of Deep Space 1, the ESA is planning to deploy a similar engine on SMART-1, which will be used over 17 months to stretch the craft’s orbit into the Weak Stability Boundary.

Advances in propulsion and miniaturization could lead to the launch of constellations of tiny craft which would communicate with one another. Testing that concept is the goal of the Space Technology 5 mission, being run from NASA’s Goddard Space Flight Center in Greenbelt, Maryland. The plan is to launch, by the end of 2003, three 21.5-kg satellites that will use micromechanical thrusters to fly in formation, maintaining a close but safe distance from one another with no input from the ground. The spacecraft will monitor the influence of solar activity on the Earth’s magnetosphere, the region of space containing charged particles trapped by our planet’s magnetic field. “We’re blazing the trail,” says Doug McLennan, who is managing the mission.

How quickly NASA, the ESA and other space agencies follow that trail remains to be seen. But if the cost of planetary missions can be reduced to a fraction of today’s hundreds of millions of dollars, it opens the door to a new kind of space science involving multiple, cheap probes — spreading the risk of failure without breaking the bank. Alkalai is even more bullish, arguing that the ‘smarter’ aspect will make failures less frequent: “Space exploration will become cheaper and more reliable.” ■

Robert Adler is a writer in Santa Rosa, California.

Web Links:

LaunchPiggy

♦ <http://www.launchpiggy.com>

Arianespace

♦ <http://www.arianespace.com>

SMART-1

♦ <http://sci.esa.int/smart>

Center for Integrated Space Microsystems

♦ <http://cism.jpl.nasa.gov>

Surrey Satellite Technology

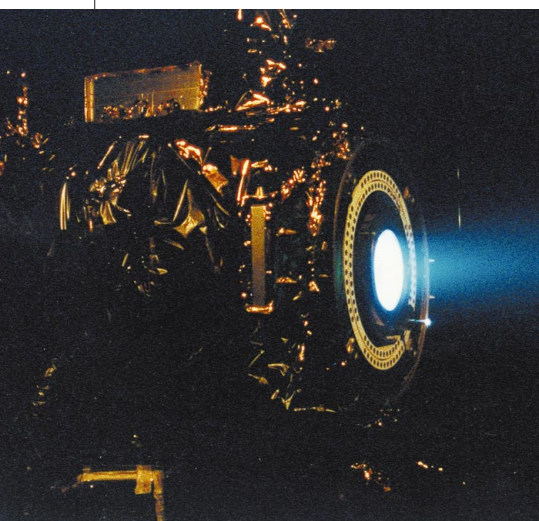
♦ <http://www.sstl.co.uk>

Deep Space 1

♦ <http://nmp.jpl.nasa.gov/ds1>

Space Technology 5

♦ <http://nmp.jpl.nasa.gov/st5/index.html>



Slow burn: the ion engine on Deep Space 1 uses a fraction of the fuel of a conventional rocket.

Important differences between sources of embryonic stem cells

Sir—Stem-cell research, including the use of human embryonic stem cells, has important implications for medical practice in the future, and the relief of suffering in many serious and intractable diseases. The European Commission's European Group on Ethics in Science and New Technologies (EGE) has recently issued an opinion on the ethical aspects of human stem-cell research. As one of the two rapporteurs of this opinion, I was disappointed by your report "European panel rejects creation of human embryos for research" (*Nature* 408, 277; 2000).

Human embryonic stem cells, which can give rise to many different cell types and tissues, are derived from early embryos *in vitro*. Your report makes little attempt to distinguish between (1) embryos donated for research by patients undergoing *in vitro* fertilization (IVF) treatment ('spare embryos'); (2) embryos made for research by fertilizing donated oocytes *in vitro* ('research embryos') and (3) embryos made for research by transfer of somatic nuclei to donated oocytes ('nuclear-transfer embryos').

The first category, spare embryos, covers all the human embryonic stem-cell lines at present being studied in the United States and Australia. The EGE concludes that in those countries (including the United Kingdom) where human embryo research is allowed, it is hard to see any specific argument that would prohibit extending the scope of such research in order to develop new treatments to cure severe diseases. Further, the group sees no argument for excluding funding of such research from the European Union's Framework programme. Thus it takes the view that, in Europe, the derivation of stem cells from IVF embryos should be not only publicly controlled but also publicly funded — unlike the situation in the United States, where state funding is not permitted to be used for this research.

The second category, research embryos, is deemed ethically unacceptable when spare embryos, including tens of thousands of frozen embryos in Europe which may become available for donation, represent a ready alternative source. The headline on your report is misleading, since the present EGE opinion deals only with stem-cell research, and does not consider the possible use of research embryos for projects that cannot by their nature be carried out on spare embryos.

For the third category, nuclear-transfer embryos, the group concludes that making

such embryos would be premature at present, in view of the extensive research still to be done on embryonic stem cells derived from spare human embryos, as well as on fetal and adult stem cells. It is this third conclusion that apparently has led you to refer to a "dramatic twist" to the debate.

Nature has an excellent reputation for clarity of presentation of scientific and social issues. It is unfortunate that in this instance, when Parliament is about to give its considered opinion on whether to extend the purposes of human embryo research in the United Kingdom in order to develop new treatments for severe diseases, an article has appeared that engenders confusion rather than clarity.

Anne McLaren

The Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, UK

Weak euro hits PhDs too

Sir—Your Careers and Recruitment report on Marie Curie fellowships (*Nature* 407, 427–429; 2000) raised the important issue of falling income among postdocs who are paid in euros but work in countries outside this currency region. We would like to draw attention also to PhD students in the same situation: they are usually less well paid than postdocs, but have similar expenses. Hence it is even more difficult for PhD students to cope with the sudden drop in value of the euro, particularly in the United States and United Kingdom.

As an example, scholarships from the Portuguese Foundation for Science and Technology (the main source of funding for Portuguese scientists) have a fixed value in euros regardless of the country in which the PhD student is working. Since the euro was introduced in January 1999, students in the United States and the United Kingdom have had to cope with a decrease in their income of 27.8% and 19.7%, respectively.

We deeply regret that the authorities, although aware of the problem, oppose fixing the value of scholarships in the currency of the country in which the student is working. If measures are not taken to correct this unfairness, the mobility of students will soon be curtailed.

Luis Graca*, Susana Nery†, Monica Bettencourt Dias‡, Tiago Magalhaes§

**Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK*
†*Skirball Institute, 540 First Avenue, New York, New York 10016, USA*

‡*Department of Biochemistry and Molecular Biology, University College London, London WC1E 6BT, UK*

§*Department of Molecular Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA*

Bernoulli was ahead of modern epidemiology

Sir—The 300th anniversary year of Swiss mathematician Daniel Bernoulli's birth is an appropriate time to reveal that he was, by a long way, the first to express the proportion of susceptible individuals of an endemic infection in terms of the force of infection and life expectancy¹.

His formula is valid for arbitrary age-dependent host mortality, in contrast to some current formulas which underestimate herd immunity (one minus the proportion of susceptible individuals), the vaccination threshold that has to be exceeded to eliminate an infection and the basic reproduction number (the inverse of the proportion of susceptible individuals)².

Bernoulli's main objective was to calculate the adjusted life table if smallpox were to be eliminated as a cause of death. He clearly defined the two epidemiological parameters, which nowadays are called the force of infection λ (the annual rate of acquiring an infection) and the case fatality c (the proportion of infections resulting in death). If L is life expectancy, then the proportion of susceptible individuals u can be written as $u = r/(\lambda L)$, where r is the cumulative incidence (the proportion of a cohort that will be infected). As the proportion c of all infected individuals die, the proportion q of all deaths due to smallpox is $q = cr$.

Hence Bernoulli estimated the susceptible proportion using the expression $u = q/(c\lambda L)$. From several large cities (especially London) which recorded cause-specific numbers of deaths, estimates of q were known to be about $1/13 = 7.7\%$. Bernoulli used Halley's life table for the city of Breslau and came up with the estimates $\lambda = 1/8$ per year and $c = 1/8 = 12.5\%$. For Paris he assumed a life expectancy of 32 years which yields a proportion of susceptible individuals of 15% ($= 107,000/700,000$), or a herd immunity of 85%.

If one assumes a constant death rate (an exponentially distributed survival time), then $u = 1/(1 + \lambda L)$ (ref. 3). This formula gives 80% as the estimate for herd immunity. The formula most frequently used today² assumes a rectangular survival function in which all individuals live until a maximum age L . Then $u \approx 1/(\lambda L)$, which gives only 75% for herd immunity. We therefore suggest that it is more accurate to use the more general formula derived by Bernoulli.

Jean le Rond d'Alembert, who had been in conflict with Bernoulli over several issues previously, violated the unwritten rules of the day by publishing a critique of Bernoulli's approach in 1761 (ref. 4), five

years before Bernoulli's essay eventually appeared in print. Bernoulli's motivation to add this formula in the final version (prepared in 1765) was explicitly to show the superiority of his 'exact' approach over the crude estimate of d'Alembert, according to whom the proportion of susceptible individuals was "much less than half".

The reason why Bernoulli's important formula has escaped notice for so long may be its cryptic presentation: one has to recover it by substituting his numerical values with their general symbols. In formulating his laudable objectives, however, he is admirably clear: "I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide."

Klaus Dietz*, J. A. P. Heesterbeek†

*Department of Medical Biometry, University of Tübingen, 72070 Tübingen, Germany

†Centre for Biometry, PO Box 16, 6700 AA Wageningen, The Netherlands

1. Bernoulli, D. *Mém. Math. Phys. Acad. R. Sci. Paris* 1–45 (1766); English translation by Bradley, L. in *Smallpox Inoculation: An Eighteenth Century Mathematical Controversy* (Adult Education Department, Nottingham, 1971).
2. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans — Dynamics and Control* (Oxford University Press, Oxford, 1991).
3. Dietz, K. *Stat. Meth. Med. Res.* 2, 23–41 (1993).
4. D'Alembert, J. *Opusculs Mathématiques*, t. II (Paris, David, 1761).

Aquaculture: part of the problem, not a solution

Sir — Aquaculture is claimed to aid the production of large quantities of low-cost protein-rich food to help feed the world, and to diminish pressure on ocean fisheries. In our opinion, neither of these claims is justified.

First, except in some parts of Asia, the main purpose of aquaculture has been to produce a luxury product for those who can afford to pay high prices. Second, Naylor *et al.*¹ analysed the consequences of aquaculture practices and indicated that the growth of global production of farmed fish and shellfish will not relieve pressure on ocean fisheries.

Naylor *et al.* also indicated that aquaculture can diminish world fisheries indirectly by habitat modification, collection of wild seedstock, food-web interaction, nutrient pollution and the introduction of exotic species. However, in this last factor the authors referred only to possible hybridization between farm escapees and wild populations of Atlantic salmon, and to the spreading of pathogens. We believe it is also necessary to consider other consequences of introducing non-indigenous organisms, such as the

elimination of autochthonous species by altering food webs, competition, hybridization and so on.

For scientists in developing countries such as ourselves, this point is of great importance. We would like to call to the attention of international organizations such as the World Bank, the International Development Bank and the Food and Agriculture Organization (FAO) the need to stop promoting aquaculture as a means of obtaining income from exports, and to concentrate on producing food for poor people. These organizations should stop promoting technological packages using exotic species and instead help the development of culture technologies for native species with the potential for aquaculture.

To give some examples: the FAO approved a technical cooperation project in Venezuela to genetically improve red tilapia, an unnecessary waste of resources that was a total failure². The International Development Bank is partially financing a programme³ to culture exotic species in Panama, including the scallops *Argopecten purpuratus*, the cachamas *Colossoma macropomus* and *Piaractus brachipomus*, the channel catfish *Ictalurus punctatus*, the Sergeant or Peacock bass cichlid *Cichla ocellaris*, the giant prawn *Macrobrachium rosenbergii* and the bullfrog *Rana catesbiana*.

The World Bank finances several programmes in Latin America (see <http://www.worldbank.org>) introducing species such as *Litopenaeus vannamei* and *L. stylirostris* into countries where they do not occur naturally. In one Honduran project, the World Bank claims there is no risk of introducing alien species because the shrimps *L. vannamei* and *L. stylirostris* occur naturally there, in the Gulf of Fonseca. But it does not mention the risk of introducing alien species in a Venezuelan project which also uses *L. vannamei*, although they are not native to that region.

Aquaculture can make unique contributions to world nutrition, thanks to its extremely high productivity in many situations and to the fact that aquatic crops are primarily protein rather than starch. Certain aquatic organisms may be better at converting primary foods than ruminants, fowl or even pigs. Some, such as filter-feeding fishes and molluscs, feed on microscopic plankton that cannot be used directly for human food.

However, if the aquaculture industry is going to reduce the pressure on wild fish stocks and provide food for the world's growing population, substantial changes must be made by governments, the private sector and international funding agencies. They must protect coastal ecosystems; promote research and development of

native species; and encourage farming of low-trophic-level fish — those low on the food chain. International technical funding agencies can exert great influence in changing practices. Otherwise, as Naylor *et al.* point out, an expanded aquaculture industry poses a threat, not only to ocean fisheries, but also to itself.

Julio E. Pérez*, Mauro Nirchio†, Juan A. Gomez‡

*Instituto Oceanográfico, Universidad de Oriente, Cumaná, Venezuela

†Escuela de Ciencias Aplicadas del Mar, Universidad de Oriente, Isla de Margarita, Venezuela

‡Centro de Ciencias del Mar y Limnología, Universidad de Panamá, Panamá

1. Naylor, R. L. *et al. Nature* **405**, 1017–1024 (2000).
2. Perez, J. E., Gomez, A. & Nirchio, M. *Interciencia* **24**, 321–323 (1999).
3. Programa de Modernización de Servicios Agropecuarios. Componente Acuicultura (Inter-American Development Bank, Panamá, 1995). <http://www.iadb.org/EXR/doc98/apr/lcpanaq.htm>

Funding would prevent waste of research time

Sir — It has become clear over the last decade that doing research in Spain is quite a heroic task (see, for example, *Nature* **407**, 428 & 941; 2000). Poor funding and lack of prospects for young researchers are among the many obstacles that have precluded the rise of Spanish science to the very top. Although an increase in the amount of money devoted to research is an obvious prerequisite, the many problems that young investigators face haven't been analysed in detail.

A numerical comparison with our EU counterparts (not to mention with the United States and Japan) shows the need to create positions to absorb many of the excellently trained postdocs that Spain has sent abroad during the past 15 years. But nobody is discussing the conditions under which these new hirings should take place. The fact is that Spain has no policy of providing start-up packages for the newly hired group leaders. This has unfortunate consequences, as new researchers sit in their offices (if they have offices) and begin chasing grants for a period of up to three years, losing all continuity with their research projects.

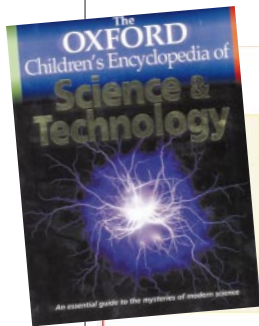
If Spain is to have a core of young investigators doing science at its best, these scientists should be provided not just with positions but also with funding for research. Otherwise, talent is wasted, again.

Pedro Martínez

Department of Anatomy and Cell Biology, University of Bergen, Aarstadveien 19, 5009 Bergen, Norway

Innovations to stir young minds

Curiosity may have killed the cat, but for tomorrow's scientists it is essential.



Internet

The Internet ('the Net' for short) is a huge computer network that connects millions of computers. It is a fast and efficient way of sending information around the world. Any two computers connected to the Internet, wherever they are, can exchange information.

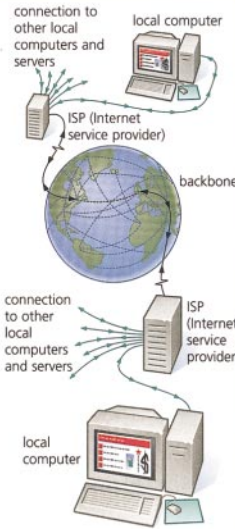
To join the Internet, you need a personal computer, a modem and a telephone line. You can use the Internet to send electronic messages to other users (this is called *e-mail*), hold electronic conversations, transfer computer files, or find information on thousands of different subjects. The number of people connected to the Internet, and the amount of information passing through it,

are increasing rapidly. Some experts think that the Net will change how we work and live.

The fastest growing area of the Internet is a system called the *World Wide Web* ('the Web' for short), which allows you to look at information stored on the Internet. To look at this information (text, pictures, sounds and video clips), you need a program called a *Web browser*. It finds the information on a *Web site*, and displays it on your screen.

The Internet works by linking together many small computer networks belonging to organizations such as businesses, universities and government departments.

► Most computers belonging to individuals, schools and small organizations that are linked to the Internet are connected through an Internet service provider (ISP).



find out more
Computers
Information technology

Howard P. Segal

Not long ago, while visiting my local public library, I found several children's books on science and technology from the 1950s, '60s and '70s. I was quickly struck by the similarities to the new works under review. Both sets of books had appealing titles, illuminating explanations of scientific discoveries and principles, clear descriptions of how tools and machines function, useful glossaries, and varying degrees of historical background, including some timelines. Regardless of the age level to which each book, old or new, was oriented, all conveyed the message that science and technology are accessible rather than mysterious. Some, although not all, further proclaimed that learning about science and technology is fun.

I am not, however, suggesting that you should patronize only used bookstores or libraries. The books discussed below not only update phenomena covered in their predecessors but also examine such relatively recent developments as personal computers, electronic communications, robotics and biotechnology.

Although the texts of all of the new books are too advanced for pre-school children, the illustrations in **Supreme Machines** and in **The Need for Speed Racing Cars** could spur interest and discussion at that level. Yet for children from ages 5 to 9, the texts of the two books, along with their illustrations, would certainly be suitable. So, too, would be **Inventions and Discoveries**, which, moreover, offer readers projects (such as making model cars, bridges, motors and

Supreme Machines: Aircraft

by Moira Butterfield

Watts: 1999. £10.99

The Need for Speed Racing Cars

by Philip Raby

Watts: 1999. £10.99

Inventions and Discoveries: The Facts Behind Amazing Breakthroughs That Changed the World

by Peter Harrison

Lorenz: 2000. £7.95

The Oxford Children's Encyclopedia of Science & Technology

Oxford University Press: 2000. £9.99

How Things Work Today

edited by Michael Wright & Mukul Patel

Marshall: 2000. £25

Wow! Inventions that Changed the World

by Philip Ardagh

Macmillan: 2000. £2.99 (pbk)

telegraphs) that reinforce the scientific and technological knowledge conveyed by the texts and illustrations.

The Oxford Children's Encyclopedia of Science & Technology and **How Things Work Today** are for children of 10 or older. Indeed, interested teenagers might find them appealing (as might adults intimidated by more technical volumes). The first two books are truly comprehensive, with many entries—from Acids to X-rays in the former, and from Urban and Domestic Environment to Space in the latter—plus an abundance of details, handsome graphics and helpful cross-references. Complementing these volumes, which offer historical

information but emphasize contemporary developments, is the slim but valuable **Wow! Inventions that Changed the World**. It provides fine mini-histories of technologies such as telephones, trains, photographs and printing-presses.

If the traditional messages that science and technology are accessible and often fun understandably persist in these contemporary books, so, too, do other, more questionable themes. First, like their predecessors, these latest works rarely distinguish science from technology. These enterprises have long been and remain largely separate, with different educational and professional requirements and different practitioners. Similarly, technology itself is broader than invention, contrary to what several of the new volumes imply. Even when, for instance, *The Oxford Children's Encyclopedia* treats science, technology and invention as separate phenomena, science nevertheless receives unjustified top billing as the alleged primary intellectual source of the other two.

Second, as with earlier books, the new ones embrace 'scientific and technological determinism', or the notion (stated explicitly in some of their titles and subtitles) that science and technology have 'changed the world' as nothing else ever has or, presumably, ever will. Contemporary historians of science and technology have repeatedly questioned this position, emphasizing instead the complex relationship between pre-existing cultural values and social structures and the varying impact on societies of scientific and technological advances.

Finally, like their predecessors, these current volumes all subscribe to the so-called 'Whig theory' of the history of science and technology. This notion derives from historian Herbert Butterfield's classic argument that English political and social history was often written as the story of unceasing progress, without significant setbacks or detours. True, *Supreme Machines* and *The Need for Speed Racing Cars* both mention vehicle crashes. And *Inventions that Changed the World* discusses a few failures, such as the Ford Edsel, and a few problems of modern cars, such as traffic, accidents and pollution. Yet the sea change throughout so much of the world in recent decades, from an unqualified belief in scientific and technological advance leading to 'progress' to a growing concern for other aspects of the 'good life'—not least, the environment—is missing from all of the works under review.

However difficult it might be for volumes such as these to enlighten children about the



suffices, especially for children. It is therefore reassuring to see that such fine books continue to be published. ■

Howard P. Segal is in the Department of History, University of Maine, Orono, Maine 04469, USA.

Spiders need friends too

Jeremy N. McNeil

In October, the first world championships for young environmental researchers brought 143 participants from 73 countries to the finals at EXPO 2000 in Hanover. WYRE — Worldwide Young Researchers for the Environment — was a joint initiative of Stiftung Jugend forscht, the organization that runs Germany's annual youth science contest, and the Deutsche Bank. It had a considerable variety of subject matter, evidenced by titles such as "Baby salamanders in danger", "Where plants still eat meat", "Day-trippers and flying mammals", "Plastics from kitchen waste", and "Making quarries green again". However, all participants had two things in

profoundly mixed blessings of science and technology, it is surely worth the effort. Are today's children not, after all, repeatedly exposed to examples of this in daily life?

Despite my reservations, these new books all deserve a wide readership. Learning about science and technology only online or solely through visits to interactive exhibits hardly

common; their dedication to improving their local environment and the belief that all individuals, including children, could make a difference.

The majority of participants indicated that their interest in environmental problems stemmed from a parent or teacher, who not only served as a resource person but also directed them to the appropriate reading material. Here are six books in the environmental field which I believe meet the need to stimulate these fertile young minds. This opinion is shared by my 12-year-old assistant Lara Cusson, whose rankings out of 10 are included at the end of each section.

Each Living Thing is an excellent book for pre-school children, underlining the importance of actually respecting all living creatures. This is very cleverly done using examples such as insects, snakes and jellyfish, species normally eliciting an 'ugh!' reaction. The illustrations are excellent and lend themselves to 'what else do you see?' questions. (9/10)

Wild and Free, also aimed at younger children, discusses species in danger of extinction, each of which is rather successfully associated with a child from the country of origin. For example, Chen the Chinese boy does not want a world without pandas and would like to see them living wild and free. The illustrations are very good, and at the end

Presents you may want to read as well

Usborne Guide to e-mail

by Mark Wallace & Philippa Wingate
Usborne: 2000. £6.99

How to Build a Robot

by Clive Gifford
Oxford University Press: 2000. £3.99

Crashing Computers

by Michael Coleman
Scholastic: 1999. £3.99

How the Future Began

by Clive Gifford
Kingfisher: 2000. £9.99, \$15.95

Terminal Chic

by Chloë Rayban
Bodley Head: 2000. £10.99

Tim Brosnan

Years ago my father bought me a train set for Christmas, which he then monopolized. It was ostensibly bought for me, but in reality for himself. I thought of this when reading the *Usborne Guide to e-mail*, which is a clear, well-written account of how to use Outlook Express and six other e-mail packages. My co-reviewers Chris and Juliette (both aged 11) read this with interest, Chris saying that it was good and that he had learnt "one or two new things, although most of the stuff we have done in school". And there is the point: it contains little that most British children will not meet at school — but a lot that

their less computer-literate parents will not know. Conclusion: pretend to buy this for your children — but read it yourself. (age 10+)

There is a brand of paint in Britain that sells itself on the slogan "Does exactly what it says on the tin". I wish this were true of *How to Build a Robot*. Chris and Juliette both picked this up, intrigued by the promise on the cover that it is a "DIY guide to building a walking, talking, thinking robot". After a few seconds browsing they discarded it. Not really surprising, as the book is a set of recycled 'home experiments' which does not begin to do what it says on the cover. Conclusion: not what it says on the tin. (age 9+)

In contrast, *Crashing Computers* was a great hit with both children. They have a collection of the same publisher's 'Horrible Histories' and 'Horrible Science', which they love, so they fought over who should read it first — Juliette won and hid it in her bedroom. I was forced to go to a local bookshop to buy another copy. Sitting there among the publisher's myriad other volumes, it was as easy to spot as a clone in the Borg Collective. It is worth the effort to dig it out — but be prepared to be assimilated to the series! *Crashing Computers* uses an engaging combination of fact and humour to tell the story of computers past, present and future — how they work, what they can do and the (often hilarious) ways humans have interacted with them.

Conclusion: entertaining, exciting and informative. Find it and buy it. (age 9+)

None of the three books mentioned so far looks like a Christmas present. They are not big or glossy and would not therefore immediately appeal to a prospective buyer. Two books that do meet these criteria are *How the Future Began* and *Terminal Chic*. They both deal with the future, but in very different ways.

How the Future Began is glossy, packed with pictures and might appeal to aunts or uncles looking for something for their nephew or niece. But would they read it? Both Chris and Juliette glanced at it, but it failed to capture their attention. I think this is because it talks at children rather than to them. This book is the antithesis of *Crashing Computers* — it is more likely to be bought but less likely to be read and enjoyed. Conclusion: would adorn a bookshelf. (age 8+)

Terminal Chic is a science-fiction book for teenage girls, telling the story of an 18-year-old transported into the year 3001. It imaginatively contrasts life now and in the year 3001, engaging the reader's imagination in a way that *How the Future Began* never does.

Conclusion: excellent, a teenage Bridget Jones transported into the far future. (age 12+) ■

Tim Brosnan is at the Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL, UK.

Each Living Thing

by Joanne Ryder, illustrated by Ashley Wolff
Harcourt. 2000. \$16, £13.95

Wild and Free

by Mike Manning & Brita Granström
Watts: 1998. £4.99

Going, Going, Gone

by Barbara Taylor
Oxford University Press: 2000. £5.99

Garbage

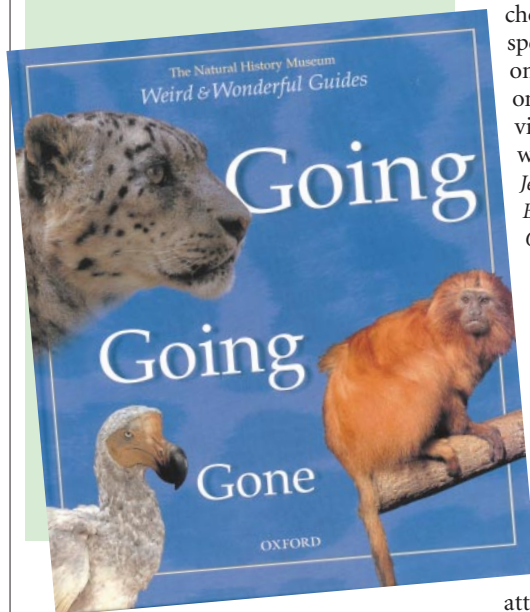
by Robert Maass
Henry Holt: 2000. \$16.95

I Want To Be An Environmentalist

by Stephanie Maze
Harcourt: 2000. \$18 (hbk), \$9 (pbk)

Planet Zoo: One Hundred Animals We Can't Afford to Lose

by Simon Barnes; illustrations by Alan Marks
Orion: 2000. £20



there is a world map showing the location of the countries discussed. (9/10)

Going, Going, Gone addresses the idea of species extinction, aimed at children of 8–12 years. This book contains a lot of good information on subjects such as the appearance of life on Earth, the idea that species have gone extinct in the past, and our ability to gain information about them from the fossil record and from species considered as living fossils. The 'did you know?' and 'true or false?' sections throughout are a definite plus and the illustrations are of high quality (8/10).

Garbage (age 12+) gives an excellent account of the accumulation of garbage and its subsequent treatment using landfills or incinerators. It then gives examples of environmentally sound alternatives for waste recycling, including practical information on composting. The accompanying photographs are really excellent (7/10).

I Want To Be An Environmentalist is one in a series of books that discuss different career paths, aimed at a teenage audience. A diversity of related issues are addressed, such

as types of environmentalist, what are the eco-issues, the history of environmentalism and famous environmentalists. This book has a strong US bias, but it contains much useful information, so I would recommend it. (9/10)

Planet Zoo: One Hundred Animals We Can't Afford to Lose (age 12+) offers a brilliant choice of species, from cuddly chimpanzees to the 'who really cares' no-eyed big-eyed wolf spider that's as bizarre as its name. Furthermore, the species occur in a diversity of habitats: marine and terrestrial, tropical to arctic. Each case makes the point that, when considering biodiversity, all species and all habitats matter. At the end are ten conservation success stories, to emphasize that we should not lose faith in our efforts to save endangered species. A nice touch is the choice of *Homo sapiens* as the hundredth species, with a reminder that we are just one species on this planet, interdependent on many others, and that our own survival is dependent on us changing our ways. (10/10)

Jeremy N. McNeil is in the Department of Biology, Laval University, Sainte-Foy, Quebec G1K 7P4, Canada.

The facts of life and death

Andrew Berry

Evaluating the science in a children's book — whether it's right or wrong — is simple. But to determine how attractive or interesting that book is to children you have somehow to acquire a child's perspective. To deal with this, I condemned 15 children's books on various aspects of animal behaviour to the hands of my brood, twins approaching four. Surprisingly, they liked the same two books that had particularly impressed me. The coincidence may, however, be due to a flaw in my experiment. Like the famous 'counting' horse Clever Hans, which picked up inadvertent cues from its owner, my offspring were perhaps sensitive to the 'you'll like this one' cues coming from their father.

The books in question are both by the writer/illustrator team of Karen Wallace and Ross Collins. **Chomp! Munch! Chew!** is "about how animals eat". The other, on animal reproduction, happily lacks a correspondingly graphic title. Both books present a lot of basic biology in an engaging and informative way — although **It Takes Two** never gets down to reproductive nuts and bolts. We are taken on a largely vertebrate-centric tour of the animal kingdom, and introduced to the snapping jaws of a gharial; the nesting habits of weaver-birds; the

nectar-harvesting devices of hummingbirds and hawkmoths; paternal care in sea horses; and the compost-incubation methods of Australian megapodes.

My reasons for liking these books are simple, and apply to children's science books in general. Most importantly, Wallace and Collins do not subscribe to the 'Guinness Book of Records' school of popular science — they present a diversity of behaviours and taxa rather than gee-whizzery. Of course, books are not the only victims of the current belief that science is no longer inherently interesting but has to be presented with highly amplified bells and whistles: many museum displays have become so dumbed down that the underlying science is drowned out in the bell/whistle cacophony. Among the books I looked at, both **I Didn't Know that Wolves Howl at the Moon** and **Zooming and Creeping** insist on addressing the pressing biological issue of what is the fastest flying moth. At least they agree that hawkmoths hold the record, although they present rather different figures, 50 and 39 kilometres per hour, respectively. Do children honestly care about lepidopteran flight velocity? I don't believe they need to view nature as a sort of collection of superheroes — fastest flyers, slickest swimmers, highest jumpers — in order to find it interesting. Perhaps more importantly, a superhero view of the natural world is as misleading as would be a view of our own species based solely on US Olympic sprinter Marion Jones, and leads to disappointment when, later in life, children encounter some real biology.



Chomp! Munch! Chew!

by Karen Wallace and Ross Collins
Franklin Watts: 1999. £10.99

It Takes Two

by Karen Wallace and Ross Collins
Franklin Watts: 1997. £4.99

I Didn't Know that Wolves Howl at the Moon

by Cecilia Fitzsimmons
Watts: 2000. £9.99

Zooming and Creeping

by Barbara Taylor
Oxford University Press: 2000. £5.99

children's books

Books these days are often elaborately illustrated with photos; this approach has been adopted by the Natural History Museum's 'Weird & Wonderful Guides', of which *Zooming and Creeping* is an example. Perhaps unexpectedly, this typically does not do animal behaviour justice. A still photo of a leaping frog is actually less expressive than a well-drawn image of the same thing; the artist can capture more than the photographer. This is certainly true of the illustrations in the Wallace/Collins books, which, although in some ways simple and cartoonish, nevertheless convey the nuances of behaviour in action much more effectively than photos.

Beatrix Potter and Walt Disney doubtless have a much greater impact on most children's appreciation of animals than do books on animal behaviour. There is therefore an inclination to anthropomorphize this kind of material, or at least to sentimentalize it. Although the Wallace/Collins books are definitely 'family viewing', they don't shy away from some of biology's less rosy aspects.

They are also keen to include humans in the natural world, and both their books incorporate our species into their survey of animal behaviour. In terms of the messages that children receive from books about biology, I think this is perhaps the most important. If children come to see humans as just another species in a richly diverse biological world, then perhaps we can look forward to the eclipse of that most absurdly persistent of all anti-scientific dogmas, creationism. ■

Andrew Berry is at the MCZ Labs, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.

Trailing nature

Sandra Knapp

Children are naturals at natural history — what parents among us have not spent hours watching a child picking up tiny stones and turning them over to examine all sides, then

solemnly putting them in the mouth to test another dimension. So it is not really a surprise that most natural history books for children are for the younger set. But surely it is those slightly older children who lose interest whom we need to keep interested in science. Perhaps publishers think the pre-teens are not interested in books at all, with this view spilling over into science and natural history books. But more and more novels are becoming available for this age group, and it should only be a matter of time before this trend affects science books as well.

Books can either be read to children or read by children themselves; the current crop has some of each. Really good books for pre-readers must have lovely illustrations and compelling words. Diane Siebert's **Cave**, an exploration in poetry of the magical underground world, is one such: any book using such words as "finite fragile works of art" to describe crystals is sure to grab at least some children's imaginations. The illustrations are lovely, as are those of **Dory Story** — a riotous adventure about messing about in boats on the east coast of the United States — with a surprising ending.

Plants are often the neglected organisms, just depicted as things for animals to sit on or eat. **Rosy Plants a Radish** is a doing-book of another kind — a lovely story about how things grow, with gardening instructions at the end. And the baobab is described in **This is the Tree** as the focus for a teeming ecosystem, all beautifully illustrated.

But the prize book (this edition is a tenth-anniversary reprint) is **The Great Kapok Tree**. It is a book with a message: the ant eater says to the sleeping man, "surely you know what happens tomorrow depends upon what you do today". The conservation ethic, not ranting but responsible, comes through loud and clear. This is a classic that in our family ranks with Dr Suess's inimitable *Lorax*.

Young school-age children (age 5–8) clamour for books that contain information, lots of it. Encyclopedias are great, both **The Usborne First Encyclopedia of Animals** and **The Oxford First Book of Animals** will satisfy a fact-hungry child. I only wish they could have been hybridized; *Usborne* has great photographs and more actual information, whereas *Oxford* treats habitats excellently and has activities that look fun.

But why do invertebrates always have to be referred to as minibeasts or creepy-crawlies? Any child who can manage *Tyrannosaurus rex* can probably manage invertebrate too. **All Kinds of Habitats** is about just that, and very good, with worldwide coverage in a slim book, and lots of activities and 'think about it' sorts of questions to stimulate an enquiring mind.

Some books are not really books, but

Empirical curiosity . . . in verse

Centrally Heated Knickers

by Michael Rosen; illustrations by Harry Horse
Puffin: 2000. £4.99

Points of View with Professor Peekaboo

by John Agard; illustrations by Satoshi Kitamura
Bodley Head: 2000. £9.99

All Sorts

by Christopher Reid; drawings by Sara Fanelli
Ondt & Gracehoper: 1999. £7.50

Maurice Riordan

Poems for children with a scientific angle are rare. There are some by Ted Hughes that have science in the background, and there are poets, such as Robert Frost and Miroslav Holub, whose scientific poems would refresh children's anthologies. But the norm in children's poetry is to appeal to fantasy and wordplay rather than to exercise an empirical curiosity about the world. So two of these books, dedicated to both poetry and science, are a new and welcome idea.

They are also, unfortunately, a disappointment. *Centrally Heated Knickers* contains 100 poems on the environment, design and technology, chemistry and physics, with 25 poems in each category. Not surprisingly, given this rigid format, the great majority are not poems at all. So the value of the book is in its exceptions, such as 'Growing Apples', which refashions the old moral of the wise servant: she counsels the king to bury an apple in order to acquire thousands more. There is verbal fun in 'Footsteps' and in 'How Humans Outdied', and some amusement in a handful of other pieces — but not enough to dispel the overall impression of an unlikely commission being fulfilled.

Points of View with Professor Peekaboo

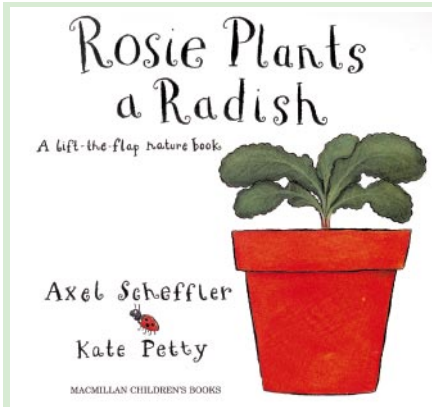
looks more promising. The eponymous professor is depicted on the cover sitting wide-eyed and pensive on top of the Earth. The drawings within are inventive and striking. The accompanying text, however, is less imaginative. Rhythm is all-important for children, but too often John Agard's lines are haphazard and the rhymes monotonous.

Let's hope that the limitations of these books do not discourage the idea of linking poetry and science. Examples of how the two work together can be found in Christopher Reid's *All Sorts*. Reid is a resourceful poet who brings to his children's verse the same standards that characterize his adult work and the poems show, in a more genial vein, the same playfulness with ideas and with language. This is as it should be.

The poems here are on mixed subjects, but some — including the best — use scientific ideas. 'A Pair of Planets', for example, creates twin alternative worlds in two 10-line verses each built on a single rhyme. The first, which is characterized by energy, rhymes hop, pop, bop, and so on; the other, where lethargy prevails, rhymes asleep, deep, heap. Another gem is 'Mud', a child's hilarious proposal to investigate the nature of mud. It's a sort of precocious grant application for "a programme of study" which, alas, cannot be undertaken "without first getting really muddy".

These poems will delight children and adults alike. And they also suggest something else: that the vein of speculative imagining, so much a part of traditional children's poetry, can be enriched by the use of science. ■

Maurice Riordan's latest poetry collection is *Floods* (Faber). He can be contacted c/o Faber & Faber, 3 Queen Square, London WC1N 3AU, UK.



Cave

by Diane Siebert
HarperCollins: 2000. \$16.95

Dory Story

by Jerry Pallotta
Charlesbridge: 2000. \$15.95

Rosie Plants a Radish

by Kate Petty; illustrations by Axel Scheffler
Macmillan: 2000. £3.99

This is the Tree

by Miriam Moss; illustrations
by Adrienne Kennaway
Frances Lincoln: 2000. £5.99

The Great Kapok Tree:

A Tale of the Amazon Rain Forest
by Lynne Cherry
Voyager/Harcourt: 2000. \$7/£5.95

The Usborne First Encyclopedia of Animals

Usborne: 2000. \$9.95, £8.99

Oxford First Book of Animals

by Barbara Taylor
Oxford University Press: 2000. £6.99 (pbk)

All Kinds of Habitats

by Sally Hewitt
Children's Press: 1999. £5.99 (pbk)

Misunderstood Bats/Misunderstood Chameleons

by Arthur John L'Hommedieu
Child's Play International: 1995. \$6.99

The Oxford Children's Encyclopedia of Plants and Animals

Oxford University Press: 1999. £9.99 (pbk)

A Leaf in Time

by David Walker;
illustrations by Mic Rolph
Portland Press: 1999. £5.25 (pbk)

Youch! It Bites:

Real-life Monsters, Up Close
by Trevor Day
Templar: 2000. £9.99, \$16

Journey to Freedom

by Colin Dann
Red Fox: 2000. £3.99

Lion Country

by Colin Dann
Hutchinson: 2000.
£10.99

Wild Things

Books:

Baboon

Rock/Leopard

Trail/Lion

Pride

by Elizabeth Laird
Macmillan:
1999/2000.

£2.99, \$2.99

Usborne

Science &

Experiments:

Ecology

by Richard

Spurgeon

Usborne:

1998.

\$8.95,

£4.99

(pbk)



toys. Pop-up books fall into this category — ours never lasted very long. But even if they don't last, **Misunderstood Bats** and **Misunderstood Chameleons** are wonderful. Both are written in a charming, matter-of-fact way for older children, while the fold-out format, with its worlds within worlds, is great for younger children. These bridge the gap, and will probably appeal to reluctant readers, as they don't really look like books at all.

Reference books about organisms are a common offering for a slightly older (9–12+) age range. **The Oxford Children's Encyclopedia of Plants and Animals** is an excellent reference book — with alphabetical entries allowing children to find information quickly. The cross-referencing is good and will encourage children to find out more for themselves. Plants and animals is a bit of a misnomer, however, as there are only a few plant entries, and none at all for mosses or ferns.

A light-hearted book about photosynthesis sounds impossible, but **A Leaf in Time** does it beautifully. The importance of plants to life on Earth is traced from the origins of life to today's need for renewable energy, all told in a way that makes the story compelling and readable. A group of 9–11-year-olds sat down and discussed photosynthesis after reading it — surely a sign that this book is doing its job.

People often assume that children prefer things that are horrible or scary — and to a large extent they are right. My 9-year-old's favourite book of all in this set was **Youch! It**

Bites, a photographic foray into the world of things that harm humans. The pictures, some close-ups, some with a scanning electron microscope, are fantastic, and the book is full of accurate information. I think the lame jokes are part of the attraction.

Novels about animals often appeal to young readers, and both Colin Dann (of 'Farthing Wood' fame) and Elizabeth Laird's stories of African adventure (**Journey to Freedom**, **Lion Country** and the 'Wild Things' series — **Lion Pride**, **Baboon Rock** and **Leopard Trail**) fall into this category. They are unexpectedly gripping, and will be enjoyed by keen animal lovers and budding natural historians alike.

Children of 12+ need books — they will still be able to use reference books such as Usborne's **Ecology**, or enjoy *A Leaf in Time*. But whereas the youngest readers are spoilt for choice in natural history, older children are largely left to fend for themselves in the world of adult books. Some can manage it easily, although these are probably the very keen who wouldn't have lost interest anyway. In today's world, where biodiversity matters to governments trying to implement the Convention on Biodiversity, knowledge about the natural world is more important than ever. It is in our own best interest to ensure that all children, not just tomorrow's taxonomists and ecologists, stay informed and interested in the world we all share. ■

Sandra Knapp is in the Department of Botany, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

.....
Dinos all around

Kevin Padian

Good popular books on dinosaurs, geared for young readers, have the same qualities as other popular science books. They're informative, accurate and entertaining. And because there are so many books on dinosaurs, it's refreshing when they're a bit different from the rest of the pack. If you're shopping for that young dino-freak on your Christmas list, you'll probably want a book that doesn't simply repeat what's already in half a dozen others by the bedside. So let's see what the elves at *Nature* have assembled for the holiday basket.

For younger readers of around 5 to 8 years old, the main attraction of Arthur L'Hommedieu's **Time Tunnel** is the layout of the book. Layered cutaways through all the pages reveal a sequence of receding murals, moving from a group of cavemen on the front cover back through the ages of the past as far as the Precambrian. The book opens out to expose a succession of pages on each geological period, complete with headlines of the major events and the usual jaw-breaking dead animal names that younger children love. Although nothing gets treated in depth, the deep-time effect is telling when the pages are all stretched out on the floor.

As a novelty concept, it is to be preferred to Anne Sharp's **Jigsaw Dinosaurs** (age 5+), which features pages of murals that become pop-out jigsaw puzzles. In contrast to chil-

children's books

dren's dinosaur books of the 1960s and 1970s, newer books have been relatively current and accurate, but this book is an exception. It is filled with statements such as "some scientists think", and "many scientists believe", which are always to be avoided (one should describe the evidence on which those inferences are based). In its crude illustrations, animals of different time periods are grouped together, a cardinal sin, and dubious speculations are treated as fact. Besides, I had trouble putting the puzzles together when more than one set of pieces came out of the book.

Also for young readers is Barbara Taylor's **Going, Going, Gone** (age 6–10), a book on extinctions that has a two-page spread on dinosaurs. Its many nice photos of recently extinct and endangered charismatic plants and animals instil awareness of what we are losing. Sadly, the book ends with a spread on rare and precious gems, but it misses out the shameful international trade in scientifically valuable specimens — living and inanimate. And the opportunity is lost to show that saving charismatic animals is not enough without saving their habitats too — a problem that probably explains the disappearance of endangered living and extinct forms. A better (but quite different) choice for this age group would be Christopher Sloan's **Feathered Dinosaurs**, which describes with authority and clarity how we know that birds are descended from dinosaurs. The book also includes the stunning evidence from new Chinese fossil discoveries of actual feathers and other such structures on dinosaurs most closely related to birds.

Joachim Oppermann's **Dinosaurs**, translated from the German, is excellent for older children (age 8+) on such matters as the derivation of the scientific names of dinosaurs and the usual basic information. It treats controversies well and explains cogently how scientists arrive at their insights, although it gets dicey in some interpretations of function and life-style. But the illustrations are not very accurate or life-like, and the text shows no evidence of consultation with scientists, which may explain some of the information given. Sue Nicholson's **Dinosaurs** is scarcely better. Its compact 'field-guide' format reduces information to boxes of factoids that mix fact with speculation.

Dinosaures: Les Seigneurs de la Terre (age 10+) is an

excellent choice for your favourite francophone. Beautiful, vivid illustrations are complemented by a cogent text that first treats many of the major ideas and issues in dinosaur science, then covers most of the best-known individual dinosaurs, their habits and vital statistics. The book has enough room to stretch out on important subjects, and so should absorb parents as well. An anglophonic version is currently in the works with National Geographic Press.

Dinosaurs: The Ultimate Guide to Prehistoric Life (age 10+) by Christopher Brochu *et al.* is every bit as good as *Dinosaures*, and perhaps even more current. There's a lot of room to explore topics, the photos and illustrations are dazzling, and the prose owes little to previous dinosaur books. A team of international experts collaborated on its text and production, and the results are impressive. Older children and adults will get as much from this informed presentation as from any dinosaur book in print. Its main deficit, surprisingly given its audience, is the lack of a rigorous explanation of cladistics (the science of reconstructing evolutionary relationships) and good coverage of how the various dinosaurs are related to each other.

For all the good points of these new efforts, the best dinosaur book ever, for my money, is David Norman's *Illustrated Encyclopedia of Dinosaurs* (Salamander, 1985) (age 10+). Although now outdated in some respects, it has the most engaging and literate prose of any

Time Tunnel

by Arthur John L'Hommedieu
Child's Play International: 2000. £2.99, \$3.99

Jigsaw Dinosaurs

by Anne Sharp
Macmillan: 2000. £9.99

Going, Going, Gone

by Barbara Taylor
Oxford University Press: 2000. £5.99

Feathered Dinosaurs

by Christopher Sloan
National Geographic Society: 2000. £11.99, \$17.95

Dinosaurs

by Joachim Oppermann
Tessloff: 2000. £7.99

Dinosaurs

by Sue Nicholson
Marshall: 2000. £4.99

Dinosaures: Les Seigneurs de la Terre

by Paul Barrett & Jose Luis Sanz
France Loisirs: 2000.
FFr122.60

Dinosaurs: The Ultimate Guide to Prehistoric Life

by Christopher A. Brochu *et al.*
Collins: 2000. £17.99



dinosaur book before or since, and incomparable illustrations by John Sibbick. Happily, Salamander has now reprinted it, along with a companion encyclopedia on flying reptiles by Peter Wellnhofer, as *The Illustrated Encyclopedia of Dinosaurs and Pterosaurs*. So it seems that the current spate of dinosaur books is more than adequate for holiday choices. ■

Kevin Padian is in the Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.

Launching kids' minds into space

Monica M. Grady

The number and variety of astronomy books aimed at children is now greater than ever. But given that astronomy is an easy subject to discuss with young children — the Sun, Moon and stars are familiar sights in the sky — it is discouraging that so few basic introductory texts have been published for a younger audience, and some of those that have are disappointing.

The Sun is my Favorite Star is a good early reading book, describing with simple language and colourful illustrations how we depend on the Sun. It is a shame that in an otherwise enjoyable book, the rainbow is shown upside down.

For 5–8-year-olds, Jacqueline Mitton and Christina Balit are very successful in their descriptions of the pictures behind constellations in **Zoo in the Sky**, a beautifully illustrated and clear description of the night sky. **Touchdown Mars!** is a fictional account of a group of children travelling to Mars, constructed so that each page is also a voyage through the alphabet. Although the book is imaginative, misleadingly the line between fact and fiction is blurred.

Non-fiction for older children tends to be quite straightforward: the four most recent astronomy texts — **DK Guide to Space**, **The Kingfisher Book of Space**, **The Usborne Complete Book of Astronomy & Space** and **Marshall mini: Space** — cover much the same aspects of the Solar System, stars and galaxies, the Universe and space exploration. All are well illustrated, the pictures being dominated by spectacular images captured by the Hubble Space Telescope and during NASA's planetary missions. All are comprehensive, and any would be a useful addition to a school library. The first three are extremely thorough, and include historical perspectives of some of the most significant astronomical discoveries, as well as maps of the night sky. *Marshall mini: Space* is more general. But, apart from placing the formation of the Solar System five billion years after the Big Bang, rather than ten bil-

Images of Earth

Lawrence W. Braile

In some ways it has never been more difficult to produce a good science book for children; in others it has never been easier. Teachers and parents, in an effort to meet higher goals associated with education reform and new science proficiency standards, are demanding books that are up to date, complete and accurate, that cover important concepts and are engaging to students. 'Less is more', as the saying goes, but nearly everyone's expectations are higher.

In this environment, it is a challenge to create a children's book that effectively describes complex science using language and illustrations that are accurate but understandable, and without 'dumbing down' the science. However, high-technology colour graphics are now available that can clearly and attractively illustrate Earth's features. **The Kingfisher Young People's Book of Planet Earth** (age 8+), for example, was runner-up in this year's Avenir children's science book prize.

Good Earth science books for children are often ones that create exciting images of the planet. These may be effective computer graphics, artwork, colourful maps, attractive and illustrative photos or interesting writing, such as eyewitness accounts of significant events or stories. **3-Dimensional Earth** (age 8+) has attractive computer-generated views of the planet mixed with informative vignettes on specific topics; it is arranged as an atlas with inset segments that highlight and describe specific features or topics. **Nature's Fury** (age 8+) uses first-person accounts to spark interest in natural disasters — an important aspect of Earth science. Like **Discovering El Niño** (ages 5–8 and 8–12), it is written in the form of stories illustrating specific Earth-science topics.

Encyclopedias such as **The Usborne**

The Sun is My Favorite Star

by Frank Asch
Harcourt: 2000. \$15, £12.95

Zoo in the Sky

by Jacqueline Mitton & Christina Balit
Francis Lincoln: 1999. £5.99

Touchdown Mars!

by Peggy Wethered, Ken Edgett & Michael Chesworth (illustrations)
Putnam: 2000. \$15.99

DK Guide to Space: A Photographic Journey Through the Universe

by Peter Bond
Dorling Kindersley: 2000. £12.99, \$15.95

The Kingfisher Book of Space

by Martin Redfern
Kingfisher: 1998. £14.99, \$21.95

Marshall mini: Space

Marshall: 2000. £4.99

The Usborne Complete Book of Astronomy & Space

by Lisa Miles & Alastair Smith
Usborne: 2000. \$8.95, £9.99 (pbk)

Journey to the Stars

by Stewart Clark
Oxford University Press: 2000. £9.99 (hbk), £6.99 (pbk)

Asteroid Impact

by Douglas Henderson
Putnam: 2000. \$16.99

How to Live on Mars: Hands-on Guide to Being a Science Superstar

by Clive Gifford
Oxford University Press: 2000. £3.99

Space Race

by Sylvia Waugh
Delacorte/Bodley Head: 2000. \$15.95/£10.99

lion years, this pocket-sized book is a handy reference text that would be invaluable as a source of images for illustrating projects.

Journey to the Stars covers much the same ground as the encyclopedias, but has the premise that the reader is on a rocket travelling to the planets, then beyond into the Galaxy, then into intergalactic space. There are frequent strange juxtapositions of Hubble images with artists' impressions of planetary and stellar landscapes which sit uneasily with the text. The hypothetical rocket journey is not really necessary and is occasionally patronizing.

Asteroid Impact is an account of the impact at the end of the Cretaceous. The combination of space, killer asteroid and dinosaurs should be a winning one, but this book does not live up to its promise. The scientific content is a straightforward account of what happened when an asteroid struck the Earth 65 million years ago. The illustrations are somewhat fanciful, especially where the dinosaurs are concerned: the picture of a *Tyrannosaurus rex* gazing up into the night sky, where an asteroid looms large over the landscape, is downright daft. **How to Live on Mars** and **Space Race** were seized by my 10-year-old son with enthusiasm. The former is one in the 'Hands-on Guides to Being a Science Superstar' series, and is an entertaining mixture of facts with simple experiments to carry

out. It is humorous, well written and copiously illustrated with cartoons. *Space Race* is fiction with no pretence at being an astronomy book, but my son found it an enthralling read. Looking through the books took us the best part of an afternoon of a wet half-term break; now he gets to read the books while I have to write about them. Sometimes life isn't fair, but I take this to be an indication that the books are hitting their target audience successfully. ■

Monica M. Grady is in the Department of Mineralogy, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

The Kingfisher Young People's Book of Planet Earth

by Martin Redfern
Kingfisher: 1999. £14.99, \$21.95

3-Dimensional Earth

by Sean Connolly
Marshall: 2000. £9.99

Nature's Fury: Eyewitness Reports of Natural Disasters

by Carole G. Vogel
Scholastic: 2000. \$16.95

Discovering El Niño: How Fable and Fact Together Help Explain the Weather

by Patricia Seibert, illustrations by Jan Davey Ellis
Millbrook: 1999. \$22.90

The Usborne Encyclopedia of Planet Earth

by Anna Claybourne, Gillian Doherty & Rebecca Treays
Usborne: 1999. £14.99

Time Life Student Library: Planet Earth

edited by Jean Burke Crawford
Time-Life Books: 1997. \$24.95

The Oxford Children's Encyclopedia of Our World

edited by Ben Dupré
Oxford University Press: 1999. £9.99 (pbk)

The Usborne First Encyclopedia of Our World

by Felicity Brooks
Usborne: 1999. £8.99.

children's books

Encyclopedia of Planet Earth (age 8+), **Time-Life Student Library: Planet Earth** (ages 8–12) and **The Oxford Children's Encyclopedia of Our World** (age 8+) include eye-catching photographs, computer-generated images and effective graphics (along with lists of Earth facts) on topics from atolls to weather. Usborne has also produced an Earth-sciences book for younger readers: **The Usborne First Encyclopedia of Our World** (ages 5–8).

Several of the books — such as *Nature's Fury* and *The Usborne Encyclopedia of Planet Earth* — provide lists of Internet addresses where you can find supplementary and recent information.

Despite the potential that exists today to produce truly excellent Earth-science books for children, many books, including those mentioned here, contain inaccuracies, introducing or reinforcing misconceptions either through the text or in diagrams. It is all too common to read that seasons change because of an annual variation in the distance of the Earth from the Sun; that the Earth's mantle is molten; that a tsunami is a tidal wave. Teachers have long observed that student misconceptions are difficult to dislodge and are a major impediment to learning. Careful review of book manuscripts and better communication between scientists and authors, editors and graphic artists are needed to improve accuracy. ■

Lawrence W. Braille is at Purdue University, West Lafayette, Indiana 47907-1397, USA.

Hurricanes and lightning

Cornelia Lüdecke

Although weather is an everyday experience, how it happens and why is a rather complicated business. Children's curiosity about it, and their endless questions, can get on parents' nerves. Mercifully, much new material is available with answers to their questions.

Two books for 5–8-year-olds — **Unser Wetter** and **The Weatherbirds** — include instructions for simple experiments demonstrating the most important aspects. For German speakers, **Unser Wetter** is very informative. **Weatherbirds**, on the other hand, tells a wonderful story of a parrot's journey from Canada back home to Costa Rica. Imaginative pictures and scientific sketches illustrate the book. In addition to the story, a basic questions-and-answers format explains the vocabulary of weather and the various phenomena encountered during the trip.

For German 10–12-year-olds, **Wetter** has been translated from the English (*Weather*; Dorling Kindersley). Its descrip-

Horribly fun and dreadfully popular!

Horrible Science Series
Horrible Geography Series

Scholastic: 1998/2000. avg 160 pp. £3.99, \$4.50

Sandra Knapp

All scientists know that science is great fun, as well as being intellectually rewarding. That sense of fun is caught brilliantly in the Horrible Science and Horrible Geography series. These series do much to fill the gap for older readers (12+), and appeal to younger children too.

The combination of jokes (in *Nasty Nature* the other fish see one with tartan skin and say "What a trendy plaice!"), with lots of interestingly presented information (Nelson's diary of the *Bounty* mutiny in *Vicious Veg* is a classic) not only entertains, but educates. Children will relish the constant reminders that they can know more than their teachers.

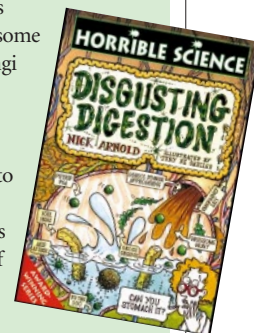
Multiple-choice questions scattered through the text make a child (or an adult!) think about what an answer might be, rather than just

learning it by rote. The pedants among us might quibble with some of the facts — for example, fungi are not really plants (maybe a sequel to *Vicious Veg: Fateful Fungi?*) — but, to be fair, the information is remarkably up to date.

One element of these books that shines out are the stories of scientists and their discoveries. Told with verve and style, the excitement of doing science really comes through. The Horrible Geography books are full of natural history: *Raging Rivers* has a section on the Amazonian fauna, and *Odious Oceans* a section about fish, crustacea and molluscs.

My daughter, when asked why she liked these books so much, said, "They are hilarious, and you learn things too." Sounds fun — just like science! ■

See page 519 for the author's address.



tions of meteorological phenomena are cogent and lucid, avoiding too much technical detail. **Wetterstation** is a German translation from the 1996 French edition (*Secrets de la Météorologie*; Gallimard Jeunesse). In addition to the historical aspects of meteorology, it gives clear descriptions of aspects such as precipitation and climatic zones. The small weather station from which the book takes its title is more of aesthetic value than anything else.

Weather & Climate, one of the books for children of 12 years and up, gives instructions for building a complete weather station using simple materials. The pages are packed with information yet remain clear. **Stormy Weather** tells exciting eyewitness stories of hurricanes and other dramatic events. Amusing sketches and cartoon stories sugar the pill of complicated correlations such as highs and lows. The book even describes events from the history of meteorology, and ends up with a look at what a modern meteorologist does.

Of the weather-tracker kits designed for children over 12 years, the best of the bunch are for German speakers. The accompanying books — which include help with weather-forecasting — are written by specialists, and **Wind und Wetter** gives clear instructions for experiments that help to show the effect of meteorological parameters. The kit includes two simple but accurate thermometers and materials for making a wooden weather-vane.

Wetterstation is the most elaborate kit. Its very professional-looking equipment includes a weather-vane and speed indicator with a small computer, and a Styrofoam weather hut. However, the scientific-looking

thermometers lack accuracy. The guide to 55 experiments is very well structured, with great emphasis on descriptions of typical cloud pictures. It is fascinating to see how you can imitate land and sea breezes, trade winds and the greenhouse effect at home. If you want a good weather book or kit, check that ready-adjusted instruments are included. Then sit back and watch your children enjoy learning about meteorology. ■

Cornelia Lüdecke is at the Institute of the History of Natural Sciences, Ludwig Maximilians University, 81371 Munich, Germany.

Unser Wetter

by Angela Weinhold
Ravensburger: 2000. DM24.80

The Weatherbirds: An Incredible Journey Through the Weather of the World

by Ted Dewan
Puffin: 2000. £5.99 (pbk)

Wetter

by John Farndon
Bertelsmann: 1999. DM10

Wetterstation

Ars-edition: 1997. DM19.90

Weather & Climate

by Fiona Watt & Francis Wilson
Usborne: 1992. £4.99

Stormy Weather

by Anita Ganeri
Scholastic: 1999. £3.99

Wind und Wetter

by Frederike Vogel & Christoph Kraul
Walter Kraul: 1996. DM47

Kachelmanns Wetterstation

by Jörg Kachelmann
& Alexander Lehmann
Franckh-KOSMOS: 1998. DM198

What dreams may come?

The scientific benefits of eating cheese before bedtime.

Paolo Mazzarello

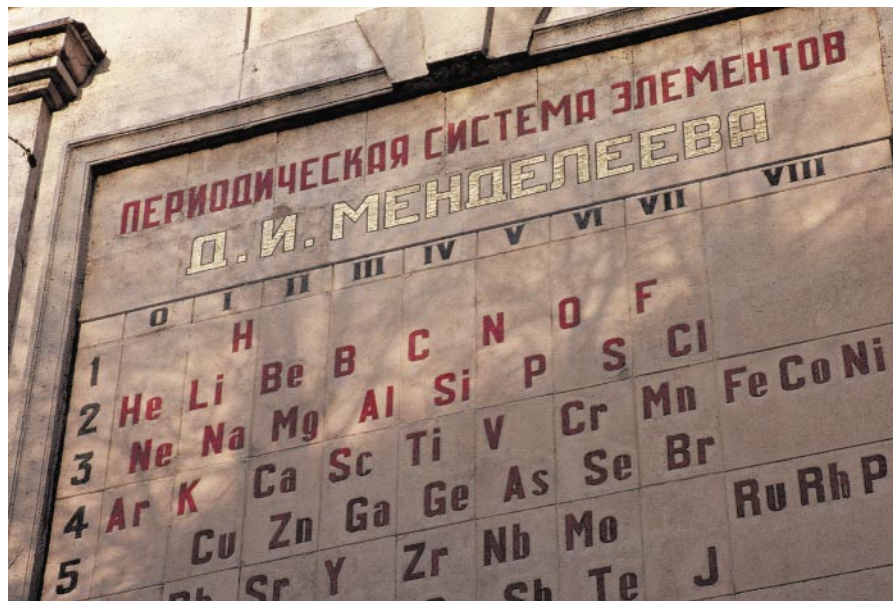
Like a snake biting its tail, the circular chain of carbon atoms danced through the dreaming mind of August Kekulé. Awakened by the vision, he spent the rest of the night developing this strange idea. Shortly afterwards he came up with a simple structure for benzene, the foundation of all aromatic substances: six carbon atoms alternately linked by three single and three double bonds to form a closed chain.

This tale of discovery inspired by a dream is not the sole example of scientific creativity catalysed by sleep. Take the autobiographical sketch by Nobel prizewinner Otto Loewi. On the night before Easter Sunday 1921: "I awoke, turned on the light, and jotted down a few notes on a tiny slip of thin paper. Then I fell asleep again. It occurred to me at six o'clock in the morning that during the night I had written down something most important, but I was unable to decipher the scrawl. The next night, at three o'clock, the idea returned. It was the design of an experiment to determine whether or not the hypothesis of chemical transmission that I had uttered 17 years ago was correct. I got up immediately, went to the laboratory, and performed a simple experiment on a frog heart according to the nocturnal design."

Loewi prepared two isolated and beating frog hearts, the first with its vagus nerve, the second without. Both were filled with a Ringer saline solution. "The vagus nerve of the first heart was stimulated for a few minutes. Then the Ringer solution that had been in the first heart during the stimulation of the vagus was transferred to the second heart." This second heartbeat slowed down in the same way as it would have after a vagus stimulation.

This result, in the words of Loewi, "proved that the nerves do not influence the heart directly but liberate from their terminals specific chemical substances which, in their turn, cause the well-known modifications of the function of the heart characteristic of the stimulation of its nerve" (see *Perspectives in Biology and Medicine* 4, 1–25; 1960).

This was the beginning of an enormous amount of research on the chemical transmission of nerve impulses that led ultimately to a new concept of brain activity. Loewi, who was at the time professor of pharmacology at the University of Graz, where he remained until his expulsion by the Nazis in 1938, realized that, while sleeping, he had begun to link the new methods he had devel-



It's all coming back to me: the periodic table was revealed to Mendeleyev in a dream.

oped to study possible substances "given off from the heart" to the hitherto unproved theory of chemical transmission.

Another example of the birth of a scientific discovery during sleep is chemistry's Magna Carta, the periodic table. In February 1869, Dmitry Mendeleyev, a 35-year-old renowned professor of chemistry at the University of St Petersburg, was puzzling over a fascinating problem: could the 63 known chemical elements be ordered by an underlying rule?

The elements could be sorted into groups according to their properties, or placed in a chain according to their atomic weights. But chemists could not uncover the relationship between chemical properties and atomic weights. During an exhausting attempt to grasp the solution — also aided by cards on which he wrote the symbols and weights of the elements, a sort of game of chemical patience — Mendeleyev fell asleep. Then: "I saw in a dream a table where all the elements fell into place as required. Awakening, I immediately wrote it down on a piece of

paper" (see *Mendeleyev's Dream* by Paul Strathern, Hamilton; 2000).

The dream had suggested an arrangement of the different groups of elements (gathered on the basis of their chemical similarities) in the order of atomic weights, allowing a regular progression with a "clear periodicity of properties" to emerge. Mendeleyev's periodic table soon became the conceptual foundation of chemical science, and immediately allowed for the prediction of hitherto unknown elements.

These three stories demonstrate that, even if rarely, scientific creativity and learning can arise during sleep. Sigmund Freud would have said that these stories are clear examples of wish-fulfilment. But a more scientific explanation is suggested by recent results obtained by Pierre Maquet *et al.* (*Nature Neuroscience* 3, 831–836; 2000). They provide evidence that memory traces are processed during REM (rapid eye movement) sleep in humans. These results seem to suggest that dreaming could be linked to a learning situation from which a new concept, only hinted at during waking experiences, could be better expressed during the dreaming reactivation and processing of memory traces, finally emerging in consciousness as a retained dream. ■

Paolo Mazzarello is at the Istituto di Genetica Biochimica ed Evoluzionistica—CNR, Via Abbiategrasso 207, 27100 Pavia, Italy.

Dreaming could be linked to the expression of new concepts only hinted at during waking.

Erratum: In the Essay "Magic beans" (*Nature* 407, 567; 2000) the picture of boats was inverted, left to right.

Too many memories

Sit down and tell me your story. I'm listening.

Pamela Sargent

I wasn't the first to practise my profession, but I was one of the pioneers, before we understood the scope of the problem we faced. My first client was Mamie Lagerfelt, and her memory files were in such disorder that she had even forgotten how many other names she had used. What she did remember very clearly (and I was struck by the irony of this later) was a news headline she had seen a century earlier — "No More Senior Moments", on a story about the first successful gene therapy for Alzheimer's disease. Mamie's great fear had always been that she would fall victim to Alzheimer's, as had several of her relatives.

"We used to call it a Senior Moment," she explained, "when you did something like forget the name of one of your children, just drew a complete blank." She no longer had to worry about Alzheimer's and having her last years filled with 'senior moments'. Her memory and her sense of herself would remain intact. You're smiling, but as I said, we didn't yet know what we were facing. We were still calling ourselves 'memory coaches' or 'personal organizers', as our clients were put off by the term 'therapist'. And back then, most of us thought that the problem was one of untidy minds that only needed straightening up.

Mamie had always been messy, and also something of a pack rat. Consequently, after having gone through her first memory-clearance procedure, she couldn't bear to get rid of anything permanently: she had downloaded and saved old memories in her personal artificial intelligence; in her town's AI; in the Novum City memory banks; in her daughter's new neural-storage system; and elsewhere, so that even her AI link couldn't locate specific memories without an extensive search — at the end of which she often forgot what memories she had been looking for in the first place. Some tidying up, proper organization of her memories in easily accessible files — I thought that would solve the problem.

It took months to help Mamie organize her memories. But it was becoming evident that her mental untidiness and exasperating incoherence were symptomatic of a deeper disorder than mere sloppiness. Colleagues reported similar problems with the clients they were coaching. "There's no framework there," Dorothea Singh said to me, "nothing to hang the memories on."

That's right — Dorothea Singh, the

founder of Narrative Therapy and Reconstruction. By now you'll have guessed that Mamie was the 'Client X' of Dorothea's most famous case history.

You already know what Dorothea's most important insight was — that the reason our client had so much trouble with her memories was that she possessed no narrative structure on which to locate them. Mamie's memories had become isolated events, fragments without context. She couldn't really remember them, even after she accessed them, because they weren't part of a story.

For Mamie, recalling too many memories was as confusing for her as not having any at all.

In retrospect, Dorothea's insight into this dysfunction seems obvious. Mamie — and other 'first-generation' longlifers — lacked a sense of continuity. They were easily distracted, incapable of extended concentration; people who had experienced life as a series of discrete moments with no unifying framework. The central principle of narrative therapy, according to Dorothea, was that memory and coherent recollection lay in the ability to tell a story, to make an epic of all of one's memories.

Mamie, as had so many others, had drifted passively from one distraction to another, and that had severely damaged her ability to weave the fabric of her memories and life into a continuous narrative. Because she couldn't tell a good story, she couldn't really remember anything, even when the incidents of her life were easily available for recollection.

Being a memory coach wasn't enough; we had to become narrative therapists, and the demand for our services could only grow. There were many like Mamie, people who had been unable to resist the distractions of their time, who had become little more than spectators of events and passive participants in simulated experiences. But we had hopes...

Yes, our failures were legion. It's almost impossible to have a successful outcome unless the client is willing to erase his past memories and start afresh. That's simply too close to death for most people: however fragmented a self might be, it rarely wants to lose that self. But in the lives of the few who had avoided Mamie's problem, we saw a remedy — better yet, a way to prevent narrative dysfunction.

These few were the fortunate ones who had exercised their minds, who had read many books in various formats and studied difficult subjects in depth, who had built up

the neural connections in their brains that gave them the continuity they needed to tell good stories. Long novels with lots of characters and narrative strands; lengthy historical accounts that enable one to understand one's historical context — that's the way to build yourself up. Work out with some Tolstoy and Balzac! Even better, keep a journal, master the art of telling a good story. All that's necessary, as the old saying goes, is to sit down and open a vein.

Excuse me — I got carried away. Historically, such pursuits were always the province of the few, and now we have even more distractions than people did in the old days. I realize that we narrative therapists are probably fighting a losing battle, but stories must be told! Our health demands this special way of ordering our minds.

Well, I hope you have a good story to tell me. I haven't heard a good one lately. ■

Nebula-award-winning Pamela Sargent's many books include the The Shore of Women (1986), Ruler of the Sky (1993) and Climb the Wind: A Novel of Another America (1999). Child of Venus will be published by Avon/Eos in 2001.



JACEY

Fighting the Ebola virus

Dennis R. Burton and Paul W. H. I. Parren

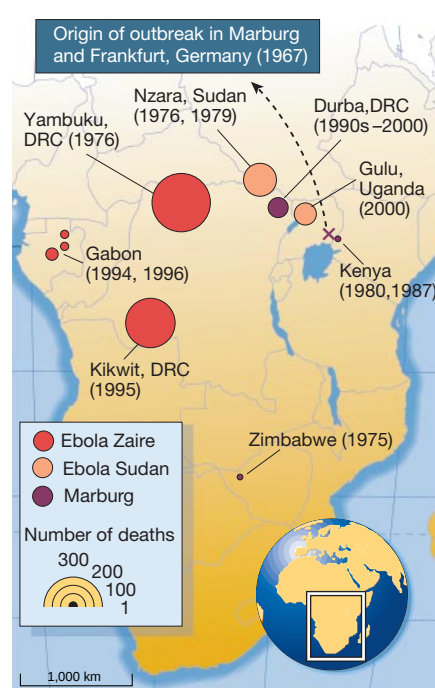
A vaccine that protects monkeys against a lethal dose of Ebola virus has been developed. But there is still a lot to learn about how this vaccine works before a version that can protect humans is available.

Perhaps no infectious agent is more associated in the modern consciousness with rapid and terrifying consequences than the Ebola virus. Regular outbreaks in Africa, together with appearances in literature and the cinema, have kept the virus in the public eye. Frequently, news reports heighten apprehension, speaking of a disease with no cure and no vaccine. Effective measures to prevent or control infection with this virus are sorely needed. On page 605 of this issue¹, Sullivan and colleagues describe a vaccine that protects monkeys against a lethal dose of Ebola virus. There's still some way to go before a human vaccine is available, but this is a step in the right direction.

Ebola virus and its lesser known cousin, Marburg virus, belong to a family called the filoviruses^{2,3}. These viruses cause haemorrhages and fevers, and are characterized by very high mortality rates, causing death in 80–90% of cases in some outbreaks. Ebola virus is transmitted through direct physical contact with infected individuals. High concentrations of virus particles are found in bodily fluids and in the skin, and the virus can be passed on through a handshake and possibly as an aerosol, through coughing and sneezing. About a week after exposure to the virus, patients become ill with influenza-like symptoms accompanied by vomiting and diarrhoea. The second phase of the illness is characterized by haemorrhages, with blood oozing from mucosal surfaces such as the gums and nose. Death usually results from shock, typically about ten days after the start of symptoms. The 'liquefaction' of internal organs sometimes attributed to the virus is fictional.

Ebola virus emerged in the African rainforest in 1976 from an unknown natural reservoir, presumed to be an animal. It caused two outbreaks at the time (Fig. 1), one in Sudan and one in the northern part of the Democratic Republic of Congo (DRC, then Zaire). The strain involved in the latter outbreak re-emerged, with an almost identical genetic make-up, nearly 20 years later some 600 miles away in the southern part of the DRC⁴. And, as we write, the Sudan strain has re-emerged to cause an outbreak in the north of Uganda⁵. Ebola virus has also caused outbreaks in Gabon and has been identified in monkeys in the Philippines.

Typically, these outbreaks flared up suddenly, probably as a result of a single trans-



mission event from the natural reservoir to a human, and were rapidly contained. Early detection of infection and isolation of patients, as well as community education, were particularly effective in halting the spread of the virus in such outbreaks⁶. But an outbreak of Marburg virus with high mortality (more than 80% of infected patients have died) has been ongoing in the Durba area in the north of the DRC for several years now, apparently involving repeated transmission of the virus from its natural reservoir to humans (S. T. Nichol and R. Swanepoel, personal communication). Such repeated transmission may represent a disturbing development in the relationship between humans and filoviruses.

Vaccination has been our most powerful antiviral strategy to date. But anyone hoping to develop a vaccine against Ebola virus has some substantial obstacles to negotiate. A vaccine that consists of a killed or attenuated (weakened) virus is unlikely, for safety reasons, to be accepted by health bodies or the public. Moreover, any vaccine must be tested under rigorous Biosafety Level 4 conditions, which require special facilities and are expensive.

Sullivan *et al.*¹ have tackled the problem by developing a new vaccine strategy. First,

they vaccinated animals with DNA encoding Ebola virus proteins (DNA immunization) — according to dogma, this type of vaccination most effectively elicits protection based on cellular immunity. Then, the immune response to these proteins was boosted with an attenuated form of a virus that normally causes colds but had been engineered to express Ebola virus proteins (adenoviral-vector immunization). Adenoviral-vector immunization is effective in inducing protective molecules called antibodies, as well as cellular immunity. Many vaccinologists think that a vaccine that elicits both types of immune response is likely to give the best results.

Having shown that their strategy was effective in protecting guinea-pigs against challenge with a rodent version of Ebola virus, the authors turned to monkeys, which can be infected directly with human strains of Ebola virus and show a clinical course similar to that seen in humans. Dramatically, whereas control animals succumbed in a matter of days following infection, vaccinated monkeys were protected and did not get sick. Moreover, three out of four of the vaccinated animals did not show any evidence of virus replication.

Does this mean that we now have in hand

a human vaccine for Ebola virus? The short answer is no, not yet. For example, Sullivan *et al.* infected the monkeys with relatively small amounts of virus, equivalent in human terms to only a few nanolitres of blood from an infected person^{7,8}. True, this amount of virus did lead to the deaths of the unvaccinated monkeys, but previous studies have shown that antibody preparations that protect against low doses of virus may be ineffective against higher doses^{9,10}. It will be crucial to know whether the vaccine strategy can protect against more substantial challenge. Also, Sullivan *et al.* did not identify the immune mechanism of protection (antibody, cellular or both), and this may be important in guiding further vaccine development. Nevertheless, coupled with earlier findings that monkeys can be protected against high doses of Marburg virus by a vaccine based on a modified alphavirus construct¹¹, it seems hopeful that human vaccination against filoviruses will be achieved.

Who will benefit from a vaccine for Ebola or Marburg viruses? The obvious answer is the local population in an outbreak area, and the medical and support personnel travelling there. In reality, however, funds are likely in the first instance to be directed towards surveillance, hygiene and barrier-nursing methods, which can be highly effective in containing an outbreak⁶. An immediate benefit of a vaccine will be to increase the margin of safety for those studying the viruses, permitting more research into the control of infection. One such aspect of research is the

search for the natural reservoirs of the filoviruses.

Finally, some have rightly raised concerns about the amount of effort spent studying Ebola and Marburg viruses, given that they affect relatively few people compared with the major pathogens in Africa, such as HIV and malaria. But our ability to predict developments in our struggle with microbes is limited. We may yet encounter more dangerous versions of the existing filoviruses, or even new ones. To be prepared, by learning how to control those viruses that are here now, is only prudent. ■

Dennis R. Burton and Paul W. H. I. Parren are in the Departments of Immunology and Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

e-mails: burton@scripps.edu

parren@scripps.edu

1. Sullivan, N. J., Sanchez, A., Rollin, P. E., Yang, Z.-y. & Nabel, G. J. *Nature* **408**, 605–609 (2000).
2. Peters, C. J., Sanchez, A., Rollin, P. E., Ksiazek, T. G. & Murphy, F. A. in *Fields Virology* (eds Fields, B. N. *et al.*) 1161–1176 (Lippincott Williams & Wilkins, Philadelphia, 1996).
3. Feldmann, H. & Klenk, H.-D. *Adv. Virus Res.* **47**, 1–52 (1996).
4. Peters, C. J. & LeDuc, J. W. (eds) *J. Infect. Dis.* **179** (suppl. 1), (1999).
5. <http://www.promedmail.org>
6. <http://www.cdc.gov/ncidod/dvrd/spb/mnpages/vhfmanual.htm>
7. Bowen, E. T. W. *et al.* in *Ebola Virus Haemorrhagic Fever* (ed. Pattyn, R. R.) 89–85 (Elsevier, Amsterdam, 1978) (available at <http://www.itg.be/ebola/>).
8. Report of an International Committee. *Bull. World Health Organ.* **56**, 271–293 (1978).
9. Kudoyarova-Zubavichene, N. M., Sergeyev, N. N., Chepurinov, A. A. & Netesov, S. V. *J. Infect. Dis.* **179** (suppl. 1), S218–S223 (1999).
10. Jahrling, P. B. *et al.* *J. Infect. Dis.* **179** (suppl. 1), S224–S234 (1999).
11. Hevey, M., Negley, D., Pushko, P., Smith, J. & Schmaljohn, A. *Virology* **251**, 28–37 (1998).

trons with tiny vibrations of the crystal lattice (phonons) leads to an attraction between the normally repulsive electrons so that they can pair up and flow without resistance.

Over the same period, experimentalists were searching for superconducting materials with higher transition temperatures, T_c , below which a material is superconducting. But progress was slow, and after a T_c of 23 K was reached in 1973, no further increase was achieved for 13 years (Fig. 1). This led to a lively discussion^{2,3} about whether 23 K was close to some theoretical upper limit to T_c .

This all changed in 1986, when superconducting copper-oxide materials were discovered⁴. The maximum T_c was quickly pushed up to well over 100 K. Unlike most earlier superconductors, the copper oxides are normally insulators rather than conductors, and charge carriers have to be introduced into the material by chemical doping before they can become superconducting. It was soon accepted that superconductivity in the copper oxides is not primarily due to a phonon mechanism, as in conventional superconductors, but to an electronic mechanism, the nature of which is still under debate. So the high transition temperatures found for the copper oxides do not rule out the possibility of a maximum T_c for phonon-driven superconductors.

Coincidentally, in 1985, chemists discovered a new form of carbon⁵ — known today as buckyballs or fullerenes. Like the copper oxides, crystalline C_{60} is normally an insulator, but can be made metallic by chemical doping. In normal C_{60} there are no charge carriers because the energy bands in its electronic structure are either completely filled or completely empty. To create a metal, the conduction band has to be partly filled with electrons (electron doping) or the valence band has to be partly emptied (hole doping). In 1991 it was found that adding alkali atoms to C_{60} crystals leads to charge transfer from the alkali atoms to C_{60} — that is, electron doping. Such alkali-doped compounds (A_3C_{60}) can become ‘metallic’⁶ and, at low temperatures, superconducting^{7,8}.

The superconductivity in A_3C_{60} is thought to be due to an interaction between electrons and phonons. The strength of this interaction is one of the factors that determine the value of T_c . Modifying the crystal

Superconductivity

C_{60} — the hole story

Olle Gunnarsson

Superconductivity has been demonstrated at a surprisingly high temperature in a C_{60} solid, raising questions about its electronic properties and hopes of even higher temperatures to come.

The sudden drop in electrical resistivity at low temperatures that characterizes superconductivity was first seen in fullerenes such as C_{60} nearly a decade ago. By introducing electrons into the carbon lattice, fullerenes can become superconducting at temperatures up to 40 K. As reported by Schön *et al.* on page 549 of this issue¹, it turns out that superconductivity can be achieved at even higher temperatures if positively charged ‘holes’ are introduced instead of electrons.

Almost a century has passed since Kamerlingh Onnes unexpectedly discovered superconductivity when he noticed that mercury’s resistivity abruptly dropped to zero at 4 K. Ever since then, this field has been an active area of research, resulting in two Nobel

prizes in 1972 and 1987. Early ideas about conventional superconductivity culminated in a theory put forward by Bardeen, Cooper and Schrieffer (BCS) in 1957. In the BCS theory, the interaction of conduction elec-

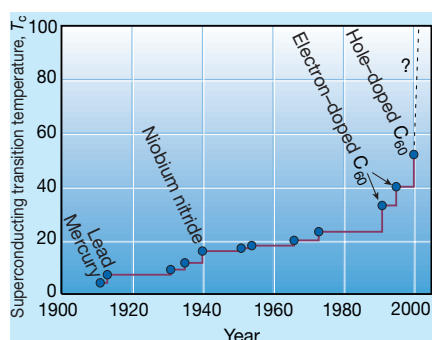


Figure 1 The maximum transition temperature, T_c , for which superconductivity has been found in conventional superconductors over the past century. It is assumed that C_{60} superconductivity conforms to this standard phonon-driven mechanism, in which tiny lattice vibrations (phonons) provide the glue that binds electrons into superconducting pairs. Electron doping of C_{60} crystals, and now hole doping of C_{60} by Schön *et al.*¹, have allowed phonon-driven superconductors to reach much higher transition temperatures.

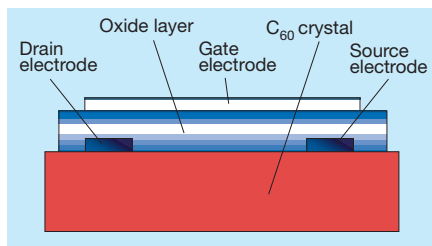


Figure 2 The field-effect transistor configuration used by Schön *et al.*^{1,11} to induce charge carriers into C_{60} crystals, changing them from insulators into superconductors.

lattice by introducing larger alkali atoms increases the electron–phonon coupling as the lattice expands. In an expanded lattice, such as Cs_3C_{60} under pressure⁹, T_c can reach 40 K. Surprisingly, Cs_3C_{60} at normal pressure is not superconducting, perhaps because of structural distortions or the closeness to a metal–insulator transition. Either way, it indicates that T_c cannot be increased further by expanding the lattice. In view of their success with electron-doped C_{60} , researchers are interested in hole-doping C_{60} , in particular because T_c is then predicted to be even higher¹⁰. But adding holes to C_{60} is very difficult, because C_{60} is strongly electro-negative, so it has not yet been achieved by chemical doping.

Schön *et al.*¹ take a completely different approach to hole-doping C_{60} . They grow an oxide layer on the surface of a C_{60} crystal and place a gate electrode on top. By adding source and drain electrodes to the underlying crystal they create a field-effect transistor¹¹ (Fig. 2). Applying a voltage to the gate electrode induces charge in the surface layer of the crystal, doping the C_{60} . If the applied voltage is positive, electrons are induced; if the voltage is negative, holes are induced. An essential aspect of this approach is the authors' ability to make devices that can take electric fields large enough to induce several electrons or holes per C_{60} molecule. They can also vary the amount of doping by simply changing the gate voltage, thereby finding the optimal level of doping for a high T_c . In earlier work they induced superconductivity at a T_c of 10 K in an electron-doped C_{60} crystal¹¹. In the present paper¹, they report a T_c of 52 K for hole-doped C_{60} .

A T_c of 52 K is very high for phonon-driven superconductivity — well above what had usually been thought possible. And with this technique, T_c may be increased even more. For electron-doped C_{60} , expansion of the crystal lattice increases T_c . With hole doping, however, T_c is already very large for the unaltered lattice structure, and its expansion, perhaps by incorporating inert molecules, could increase T_c substantially. Just as with electron-doped C_{60} , this approach would probably also fail for hole-doped systems if the lattice is expanded too much — for instance, the system could then become

an insulator. But the ability to continuously change the doping, even to fractional numbers of charges per molecule, may allow T_c to be increased substantially before this happens.

This work opens up several interesting avenues. The possibility of comparing electron- and hole-doped C_{60} , and of varying the doping continuously, should improve our understanding of the electronic structure in C_{60} in general, and of superconductivity in C_{60} in particular. With Schön *et al.*'s field-effect transistor, only the outermost layer of C_{60} molecules is thought to be doped — the system is said to be quasi two-dimensional (2D). It will be interesting to compare this situation with chemically doped 3D systems, both in terms of the electronic behaviour and in terms of 2D versus 3D superconductivity. There may also be potential for making practical devices, because with this system

it is possible to switch C_{60} between insulating and superconducting behaviour. For instance, one can imagine it being used as an ideal switch¹¹.

Olle Gunnarsson is at the Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany. e-mail:gunnar@and.mpi-stuttgart.mpg.de

1. Schön, J. H., Kloc, Ch. & Batlogg, B. *Nature* **408**, 549–552 (2000).
2. Cohen, M. L. & Anderson, P. W. in *Superconductivity in d- and f-band Metals* (ed. Douglass, D. H.) 17–27 (AIP, New York, 1972).
3. Dolgov, O. V., Kirzhnits, D. A. & Maksimov, E. G. *Rev. Mod. Phys.* **53**, 81–93 (1981).
4. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
5. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
6. Haddon, R. C. *et al. Nature* **350**, 320–322 (1991).
7. Hebard, A. F. *et al. Nature* **350**, 600–601 (1991).
8. Tanigaki, K. *et al. Nature* **352**, 222–223 (1991).
9. Palstra, T. T. M. *et al. Solid State Commun.* **93**, 327–330 (1995).
10. Mazin, I. I. *et al. Phys. Rev. B* **45**, 5114–5117 (1992).
11. Schön, J. H., Kloc, Ch., Haddon, R. C. & Batlogg, B. *Science* **288**, 656–658 (2000).

Evolutionary biology

The problem of variation

David L. Stern

One genetic source of the sex-specific variation in pigmentation patterns of different fruitfly species has been identified. This study illustrates the power of bringing together developmental and evolutionary biology.

Put simply, evolution is the result of natural selection promoting or eliminating particular heritable 'phenotypic' differences in a population, such as a behaviour or a physical characteristic. Over the past 150 years, many aspects of this process have been explored. But one factor in the equation has remained enigmatic — the source of the heritable phenotypic variation itself.

On page 553 of this issue¹, Kopp and colleagues illustrate how developmental genetics can help to solve this problem. Their first finding is interesting enough: they have identified a key genetic regulator of sex-specific differences in abdominal pigmentation in the fruitfly *Drosophila melanogaster*. Even more impressive, they show that changes in the regulation of this gene may have contributed to the variety of pigmentation patterns seen among different fruitfly species. Finally, they investigate how these findings relate to mating behaviour.

Nowadays it is thought that mutations in DNA sequences cause changes in phenotype. But it has proved difficult to identify evolutionarily relevant mutations — those that alter phenotypes and persist in natural populations. There are three core difficulties. First, the DNA regions that encode genes contain vast quantities of so-called neutral variations, which have no effect on either the phenotype or the fitness — the ability to survive and reproduce — of an organism.

Second, many evolutionarily relevant mutations may occur outside the protein-coding regions of genes. Genes can be divided roughly into the DNA sequence that codes for RNA, which may in turn be translated into protein, and regulatory regions, which encode instructions for regulating gene expression. These instructions are deciphered by DNA-binding proteins called transcription factors, which recognize specific DNA motifs and enhance or repress transcription. In doing so, they switch genes on or off in specific spatial and temporal patterns throughout development². Although regulatory regions have a critical role in development, we still have a poor understanding of how their structure relates to their function. The third difficulty is that even dramatic evolutionary alterations in the DNA of regulatory regions can result in no change in function^{3,4}.

The upshot of all this is that many evolutionarily relevant changes in DNA sequences are probably buried within vast quantities of neutral variation, within both protein-coding sequences and poorly understood regulatory regions. Another obstacle to identifying evolutionarily relevant mutations is the paucity of experimental systems that allow physical variations to be related to DNA variations.

The new discipline of evolutionary developmental biology is ideally poised to tackle these problems, by applying the experimen-

tal approaches of developmental biology to the study of variation between closely related species. Developmental biologists typically work out the normal function of genes by examining the physical consequences of experimentally induced mutations within those genes. Evolutionary geneticists, by comparison, focus on the natural genetic variation found within and between species. Natural genetic variation normally causes more subtle phenotypic differences than the mutations studied in developmental biology. But it is becoming clear that researchers in these two fields have been studying different kinds of variation at the same genes^{5–8}.

The work of Kopp *et al.*¹ illustrates how developmental biology can help to identify the source of heritable phenotypic variations. The authors have used a model experimental organism — *D. melanogaster* — to gain a detailed genetic understanding of a phenotypic trait that varies between fruitfly species, and have determined whether variations in this gene between species correlate with phenotypic variations.

Kopp *et al.* first show that differential regulation of the *bric-a-brac* (*bab*) gene, which codes for a putative transcription factor, is required for the development of the sex-specific abdominal pigmentation patterns in *D. melanogaster*. In females, *bab* is expressed in all abdominal segments, repressing pigmentation. In males, *bab* is not expressed in the most posterior segments, leading to pigmented posterior abdomens (see Fig. 1 on page 553). The authors discovered that the *bab* gene integrates inputs from the sex-determination pathway, which limits repression of *bab* to males, and from the Hox genes, which limit *bab* repression to the posterior abdominal segments.

The authors then show that this sexually 'dimorphic' regulation of *bab* sometimes correlates with sex-specific pigmentation patterns in other *Drosophila* species. The implication is that the evolution of the regulation of *bab* was important in the evolution of this sexually dimorphic feature. However, in some species of the *montium* subgroup, *bab* is repressed in the posterior abdominal segments of males even though the species do not have sex-specific pigmentation. So, the genetic circuitry linking *bab* to pigmentation has also evolved, presumably as a result of alterations in the regulatory regions of genes downstream of *bab*. This link was probably either lost in the *montium* subgroup, or gained in the other subgroups.

Kopp *et al.* also examined other sexually dimorphic characteristics in *D. melanogaster*, such as the shape of abdominal segments and the patterns of bristles and of small projections of the cuticle called trichomes. They find that sexual dimorphism in these features also depends on the regulation of *bab*. In addition, in those species most closely related to *D. melanogaster*, male-specific expression of these characters correlates with the repression of *bab*. In contrast, species of the *ananassae* and *montium* subgroups have diverse bristle and trichome patterns that vary independently of *bab* expression patterns. In these lineages, there may have been evolutionary changes in unidentified genes that in *D. melanogaster* are regulated by *bab*, the sex-determination pathway and the Hox genes.

Finally, the authors have investigated the role of pigmentation patterns in sexual behaviour. They find that *D. melanogaster* females do not prefer males with a normal pigmentation pattern over males with a

mutant, unpigmented posterior abdomen. Males, however, do discriminate against females with a 'male' pigmentation pattern. So males — but not females — may use pigmentation cues to discriminate between the sexes. Kopp *et al.*'s use of mutations within *D. melanogaster* to study sexual behaviour is innovative, and potentially links selection acting on mating systems to the molecular mechanisms underlying phenotypic evolution. But it is not yet clear whether these results are relevant to sexual behaviour in flies with natural variations in pigmentation.

This difference between the variation generated in the lab and that seen in natural populations illustrates one historical difference between developmental and evolutionary biology⁹. But Kopp *et al.*'s work¹ shows how these two fields can be brought closer together. No doubt more progress will come from further recognition that both fields are interested in studying the same genes, and from the identification of the specific genetic changes that cause natural phenotypic variations. ■

David L. Stern is at the Laboratory for Development and Evolution, University Museum of Zoology, Downing Street, Cambridge CB2 3EJ, UK.
e-mail: ds205@cus.cam.ac.uk

1. Kopp, A., Duncan, I. & Carroll, S. B. *Nature* **408**, 553–559 (2000).
2. Arnone, M. I. & Davidson, E. H. *Development* **124**, 1851–1864 (1997).
3. Ludwig, M. Z., Bergman, C., Patel, N. H. & Kreitman, M. *Nature* **403**, 564–567 (2000).
4. Takahashi, H., Mitani, Y., Satoh, G. & Satoh, N. *Development* **126**, 3725–3734 (1999).
5. Gibson, G. & Hogness, D. S. *Science* **271**, 200–203 (1996).
6. Stern, D. L. *Nature* **396**, 463–466 (1998).
7. Long, A. T., Lyman, R. F., Morgan, A. H., Langley, C. H. & Mackay, T. F. C. *Genetics* **154**, 1255–1269 (2000).
8. Sucena, E. & Stern, D. L. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).
9. Stern, D. L. *Evolution* **54**, 1079–1091 (2000).

Nanotechnology

Pinning on impact

When it comes to investigating phenomena at the nanometre scale, clusters are all the rage. These are minute particles, containing anything from tens of atoms to a few thousand, which exhibit size-dependent electrical, optical, catalytic and mechanical properties.

To study the behaviour of clusters, so-called 'nanoislands' — isolated clusters of nanometre dimensions on a supporting substrate — are often used. One approach is to form the clusters in a gas phase and then deposit them on the substrate. The trouble is getting the resulting nanoislands to stick. R. E. Palmer and colleagues have looked into the problem of

pinning them down and have devised a system that involves smashing clusters of silver into a graphite substrate (*J. Chem. Phys.* **113**, 7723–7727; 2000).

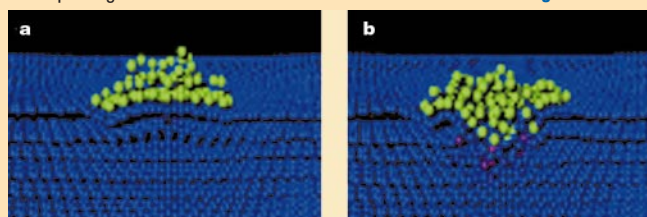
Palmer's group use a beam of charged and size-selected silver clusters, accelerated in an electric field and directed onto the graphite surface. The simulation snapshots shown here illustrate the process for a silver cluster containing 147 atoms. When accelerated in an electric field to an energy of 1,500 electron volts (**a**), the cluster simply flattens on impact without penetrating the graphite surface. But at 2,000 electron volts (**b**), the cluster disrupts the carbon lattice of the graphite, and silver atoms

implant into the surface. The onset of pinning — defined as one or more carbon atoms being displaced from their lattice site — occurs between 1,625 and 1,750 electron volts.

Similar behaviour is observed with clusters of 50 to 200 atoms in both molecular dynamics simulations and experiments using scanning tunnelling microscopy. The pinning threshold increases

linearly with cluster size, suggesting that the energy required to remove one carbon atom from the lattice is transferred in the initial elastic collision. Pinning through the creation of defects beneath the incident cluster is particularly effective in this system; whether the approach might be applied to creating practically useful nanostructures remains to be seen.

Magdalena Helmer





100 YEARS AGO

Huxley's life and work II. "When I reached intellectual maturity," Huxley tells us, "and began to ask myself whether I was an atheist, a theist or a pantheist, a materialist or an idealist, a Christian or a freethinker, I found that the more I learned and reflected, the less ready was the answer; until, at last, I came to the conclusion that I had neither art nor part with any of these denominations except the last"... These considerations pressed forcibly on him when he joined the Metaphysical Society. "Every variety," he says, "of philosophical and theological opinion was represented there, and expressed itself with entire openness; most of my colleagues were 'ists' of one sort or another; and, however kind and friendly they might be, I, the man without a rag of a habit to cover himself with, could not fail to have some of the uneasy feelings which must have beset the historical fox when, after leaving the trap, in which his tail remained, he presented himself to his normally elongated companions. So I took thought, and invented what I conceived to be the appropriate title of agnostic. It came into my head as suggestively antithetic to the gnostic of Church history, who professed to know so much about the very things of which I was ignorant; and I took the earliest opportunity of parading it at our Society, to show that I, too, had a tail like the other foxes."

From *Nature* 29 November 1900.

50 YEARS AGO

The Rumford Medal is awarded to Sir Frank Whittle for his inventions relating to the production and application of power... He proposed to take all the air that was necessary for the propulsion of aircraft through a compressor, raise it to a high temperature by the combustion of fuel and expand it through a turbine. The expansion provides all the energy for the operation of the compressor, and the kinetic energy in the exhaust gases provides the propulsive effect... As appears to be the case with many pioneers of mechanical engineering, brilliance of inventive genius is coupled in him with great powers of drive, leadership and organization. Whittle is a worthy successor of Watt and Parsons. Thanks to him, the magic carpet is no longer a fantasy of an Arabian Nights Entertainment.

From *Nature* 2 December 1950.

Biogeography

Chile refuges

Peter D. Moore

Where did warmth-loving species shelter from the cold during ice ages? Genetic analyses of a coniferous tree in South America provide an unexpected answer.

The past two million years have been tough for species that thrive in temperate conditions. This period, the Pleistocene epoch, has seen erratic, often rapid climatic fluctuations that have tested many species to their limits, and some beyond. The persistence of some temperate species into the present day indicates that they must have survived the trying climatic conditions — but where? The identification of the precise locations of glacial survival could prove important in understanding current patterns of species distribution.

This work has traditionally relied on the efforts of geologists and palaeontologists, but genetic analyses of present-day populations are proving a valuable source of further information. This kind of approach has been used in the Northern Hemisphere to study species ranging from lodgepole pines¹ to woodrats². Writing in the *Journal of Biogeography*³, Premoli and colleagues use genetic analyses to study the coniferous tree *Fitzroya cupressoides* (Fig. 1), and manage to unravel the glacial history of this species in South America.

The climatic shifts of the past two million years have selected in favour of those species that can adjust their patterns of distribution, moving from one place to another and persisting — often in suboptimal conditions — in isolated refugia during cold periods. Reliably dated fossils have provided clues to the locations of these refugia and to subsequent patterns of spread during warm episodes, but fossils do not always supply conclusive evidence. In addition, the resolution of radiocarbon dating at the start of the most recent interglacial period, some 10,000 years ago, is poor. This makes it difficult to determine precise patterns of species movements. Pollen analyses have also been used extensively to identify refugia, but these studies suffer from problems of interpretation. For example, how much pollen needs to be found to assure us of a plant's presence in an area?

The conditions in the southern part of South America provide a good example of these types of problem. The Andes mountain chain separating Argentina and Chile was glaciated during the last glacial maximum, around 22,000 years ago, while the western lowland part of Chile was ice-free. Pollen studies have indicated that these lowlands were occupied by trees throughout the glaciation, so this region probably acted



Figure 1 *Fitzroya cupressoides*, a long-lived South American conifer. Genetic analyses of living *Fitzroya* populations³ provide clues to the location of refuges for temperate organisms in the Southern Hemisphere during the last ice age.

as a refugium for temperate forest organisms. But forest with a similar composition has also developed on the eastern side of the Andes during the present-day interglacial period, so refugia may have existed there, too.

To investigate this question, Premoli *et al.*³ have taken a genetic approach, based on the argument that patterns of genetic diversity among living populations of a species should provide insight into its history. They studied the alerce, or Patagonian cypress (*Fitzroya*), which occurs today in lowland Chile, on the western slopes of the Andes, and in scattered populations on the Argentinian (eastern) side of the mountains. *Fitzroya* is a long-lived conifer that has been recorded to survive for over 3,600 years, a lifespan second only to that of the bristlecone pine. Current regeneration of the tree is poor, suggesting that it would not have been able to spread rapidly after the end of the last glaciation.

Premoli *et al.*³ studied 24 populations of *Fitzroya* from both sides of the Andes, looking at 11 enzymatic systems involving 21 putative genetic loci. This provided an

GALEN ROWELL/CORBIS

adequate means of estimating genetic variability within the species. If the western side of the Andes constituted the only refugial area, then modern-day populations would be expected to have overall genetic similarity. One would also expect lower diversity in populations found in the east, because these would have been established by a limited number of individuals that managed to disperse across the ice-laden divide of the mountain chain. But if there were separate eastern and western refugia, distinct genetic differences would be expected between today's eastern and western populations, because they will have been isolated from each other since at least the start of the last glaciation.

The authors found that the eastern and western populations are genetically distinct, so it seems that the single-refugium hypothesis must be abandoned. On the face of it, it is surprising that warmth-demanding species survived in the interior of the South American continent, which one might expect to be colder than the coasts. Perhaps areas east of the Andes had relatively little snowfall (because of the precipitation shadow effect), limiting ice development and leaving some sites ice-free.

The genetic patterns of the eastern populations are also diverse, suggesting that there

were many refugia on the Argentinian side of the Andes. The more southerly of the eastern populations are the most varied, implying that those in the northeast were derived from those in the south as a result of postglacial spread. This is also surprising: one would expect refugia for temperate trees to be found at warmer latitudes further north. Premoli *et al.* question whether the extensive latitudinal movements seen widely in the Northern Hemisphere are to be expected in the Southern Hemisphere, where very different climatic patterns and less extreme variations were experienced in the Pleistocene. It seems that the biotic impact of the glacial cycles, in terms of both the latitudinal extent of vegetation shifts and the frequency of extinctions, was greater in the Northern Hemisphere. For Northern Hemisphere biogeographers, the south is clearly a foreign country and plants do things differently there. ■

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 8WA, UK.
e-mail: peter.moore@kcl.ac.uk

1. Cwynar, L. C. & MacDonald, G. M. *Am. Nat.* **129**, 463–469 (1987).
2. Smith, F. A., Matocq, M. D., Melendez, K. F., Ditto, A. M. & Kelly, P. A. *J. Biogeogr.* **27**, 483–497 (2000).
3. Premoli, A. C., Kitzberger, T. & Veblen, T. T. *J. Biogeogr.* **27**, 251–260 (2000).

Earthquake science

Shaking faults loose

Chris Marone

Earthquakes often induce aftershocks on other faults, but the mechanisms remain elusive. An innovative analysis tells us more about the effects of dynamic stresses caused by the passage of seismic waves.

How does an earthquake on one fault affect the likelihood of failure on nearby faults? In tackling this question, Kilb *et al.*¹ (page 570 of this issue) distinguish between two different triggering effects and document the possible influence of one of them — seismic waves.

For anyone who has ever sat down with a child to play *Don't Spill The Beans* — a game in which players take turns adding beans to a shallow pot that pivots like a see-saw — it will be no surprise that earthquake triggering depends on some details of the initial earthquake rupture. In *Don't Spill The Beans*, dropping a bean in the pot is more likely to cause a spill than carefully adding one over the pot's pivot, even though the static load is the same in both cases. Likewise, if fault slip occurs slowly and without dynamic shaking (add that bean cautiously) the question is whether the static stress transferred to the surroundings is enough to cause failure on neighbouring faults. But earthquakes also produce seismic waves which cause intense

shaking in the immediate region of the fault concerned. These dynamic stresses might also trigger aftershocks — rather like bumping the pot.

A key problem in earthquake triggering is that of understanding the effects of static stress transfer compared with those of dynamic stresses and shaking-induced fault weakening. Until now, distinguishing between these effects has been possible only when the secondary earthquake is triggered far from the original². In this situation, changes in static stress are negligible; only dynamic stresses are significant. But 'directivity', which can amplify shaking in the direction of earthquake rupture, provides a way to distinguish between dynamic and static stresses close to an earthquake. Kilb and co-workers¹ apply this approach to the magnitude 7.3 Landers earthquake of 1992. They find that aftershocks are more likely in areas of high dynamic shaking — as long as the change in static stress does not have the opposite effect and inhibit fault failure.

Research on earthquake triggering and fault interaction has been driven in part by the need to understand changes in seismic hazard on human timescales. The situation is particularly clear in the San Francisco Bay area (Fig. 1, overleaf): a large earthquake on either the San Andreas or the Hayward fault could increase or decrease seismic risk across the entire area. But how do we codify fault interaction, and what causes one earthquake to trigger another?

One factor is surely the transfer of static (ambient) stresses^{3,4}. Earthquakes release the stresses supported by faults and, because earthquakes are shear failures — slip parallel to the fault dominates over any net perpendicular (normal) motion — they primarily alter the shear stresses parallel to the fault. But on some faults both the normal and shear stresses are affected. Fault strength is a friction problem, so understanding earthquake triggering by stress transfer involves evaluating whether the change in so-called Coulomb failure stress (shear stress plus normal stress multiplied by the coefficient of friction) is enough to cause fault failure or inhibit it. Close to a rupture, where the potential for damaging aftershocks is greatest, the change in static stress may be of the same order as the amplitude of dynamic shaking (see Fig. 1 of the paper on page 570). So in this circumstance, aftershocks could be the result of changes in static stresses, dynamic stresses or both.

Kilb and co-workers¹ show how the question can be addressed in the immediate vicinity of large earthquakes. They also introduce a new wrinkle by explicitly acknowledging that the frictional properties of faults are likely to be altered by dynamic shaking. Their thinking is sound — we have all seen stubborn jar lids come off after a good whack against the kitchen counter — but the principle has been difficult to test with previous methods and data.

In theory, the effects of dynamic stresses carried by seismic waves can be evaluated using the Coulomb failure stress and the approach developed for static stresses. But two difficulties arise.

First, the most sensitive way to evaluate fault interactions is to measure changes in the rate of seismicity. Ideally, one would evaluate variations in seismicity rate as a function of time, in the form of Fig. 1 of Kilb and co-workers, and compare them directly with the dynamic (time varying) and static Coulomb failure stress. But in most cases it is not possible to measure seismicity rates with sufficient temporal resolution. Second, we do not expect instantaneous failure of another fault, even when the stresses exceed a nominal threshold. Modern friction laws⁵ show that brittle frictional strength depends on strain-rate and slip history in a way that produces delayed failure in many cases. The effects are small, but

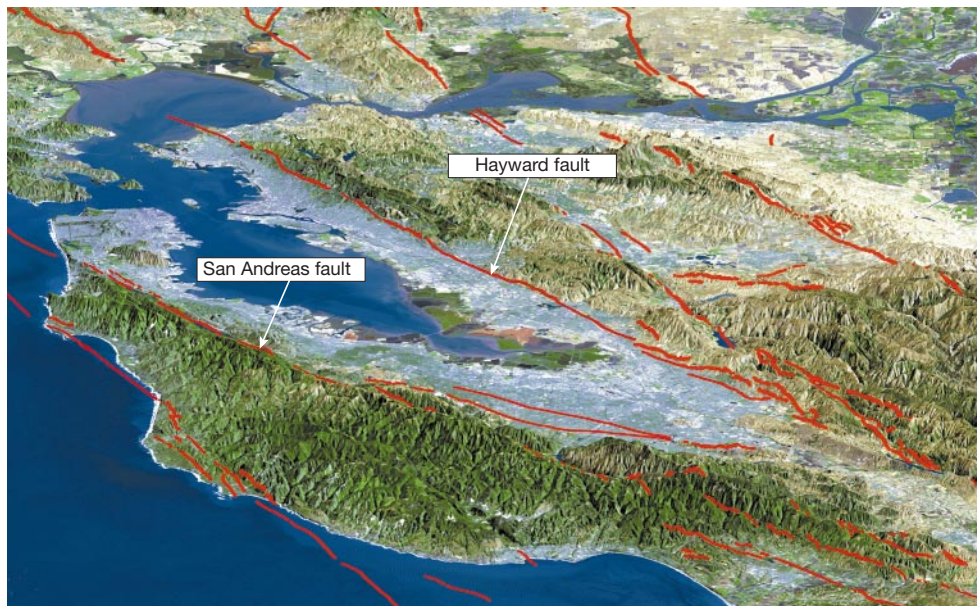


Figure 1 The San Francisco Bay area is cut by several large faults, including the San Andreas, which runs through the city of San Francisco, and the Hayward fault east of the bay. A large earthquake on one fault could trigger earthquakes on other nearby faults, but the mechanisms of such interactions are poorly understood. Possible triggering agents include transfer of static stresses, dynamic stressing by seismic waves, or shaking-induced fault weakening; the latter two in particular are addressed by Kilb *et al.*¹. Distinguishing between these mechanisms is important for understanding earthquake-induced changes in seismic hazard. (Figure modified from ref. 3.)

critical to the initiation of earthquake-like instabilities⁶.

Kilb *et al.* have taken advantage of strong rupture directivity in the Landers earthquake to distinguish between dynamic and static stresses. Directivity effects become significant when the speed of rupture propagation approaches that of radiated seismic waves. Dynamic stress amplitudes are increased in the propagation direction, but static stresses are unaffected. The Landers mainshock propagated from southeast to northwest, producing much greater shaking to the northwest. Kilb *et al.* show that this region of high dynamic stressing extended at least 100 km from the end of the mainshock, producing large differences in the spatial pattern of static and dynamic stresses, and identifiable aftershocks to the northwest.

In the same way that increases in Coulomb stress promote fault failure, stress decreases should inhibit failure, and one factor that Kilb and co-workers do not address is such 'stress shadowing'³. Proponents of static-stress-transfer hypotheses for fault interaction point out that neither dynamic stress transfer nor fault-weakening by dynamic shaking can explain stress shadows. This could seem a criticism of Kilb *et al.*'s approach. But to my mind it is unlikely that dynamic stressing would produce stress shadows in any case: dynamic shaking is likely to strengthen faults only under unusual conditions. One way to prove the role of dynamic weakening would be to document a shaking-induced increase in seismicity within a static stress shadow, but this would be difficult given present data limitations.

Distinguishing between the influence of static and dynamic stresses in earthquake triggering is intriguing, and may help in building a more general understanding of fault interaction. In the future, initiatives

such as EarthScope⁷ should allow detailed study of surface deformation and earthquake-induced strain changes in the Earth's crust. Those data, and more work of the sort reported by Kilb *et al.*, will shed further light on earthquake triggering and the associated changes in seismic hazard. ■

Chris Marone is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts

Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

e-mail: cjm@westerly.mit.edu

1. Kilb, D., Gomberg, J. & Bodin, P. *Nature* **408**, 570–574 (2000).
2. Hill, D. P. *et al.* *Science* **260**, 1617–1623 (1993).
3. Stein, R. S. *Nature* **402**, 605–609 (1999).
4. Harris, R. A. *J. Geophys. Res.* **103**, 24347–24358 (1998).
5. Marone, C. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
6. Scholz, C. H. *Nature* **391**, 37–42 (1998).
7. <http://www.earthscope.org>

Osteoimmunology

Bone versus immune system

Joseph R. Arron and Yongwon Choi

A molecule on activated T cells triggers bone loss by switching on bone-resorbing cells. Fortunately, it seems that this mechanism is kept in check by another molecule, secreted by the T cells.

The importance of bone to the immune system is well known: immune cells form in the bone marrow. But the importance of immune cells to bone is less clear. Little is known beyond the fact that normal bone growth and restructuring are disrupted in disorders such as autoimmune diseases. Several years ago, however, a particular molecule on the surface of activated T cells was found to activate bone-resorbing cells¹. This raised a problem: T cells are working constantly to fight off the universe of foreign particles in which we live, so, at any point in time, some T cells are activated. What prevents these T cells from causing extensive bone loss? On page 600 of this issue, Takayanagi and colleagues² provide an answer: T cells also secrete a molecule that inhibits the development and activation of the bone-resorbing cells. This finding underscores the dynamic relationship between bone and immune system.

Often thought of as a rigid, unchanging

entity, skeletal bone is actually the result of a dynamic process involving the secretion and resorption of the bone matrix. These opposing actions are carried out by two cell types — osteoblasts and osteoclasts, respectively — and must be kept in balance to maintain skeletal integrity and calcium metabolism, as bone is the main source and repository of the body's calcium. The most common problem with this balancing act occurs when the rate of resorption exceeds the rate of mineral deposition. This results in a loss of bone mass, as seen in osteoporosis, many inflammatory diseases, such as rheumatoid arthritis, and many cancers³. One way of maintaining the balance is to keep in check the resorption of bone by osteoclasts.

Over the past few years, the molecular mechanisms underlying the maturation and activation of osteoclasts have been worked out from studies of genetic alterations resulting in skeletal abnormalities in mice, rats and humans. Osteoclasts are

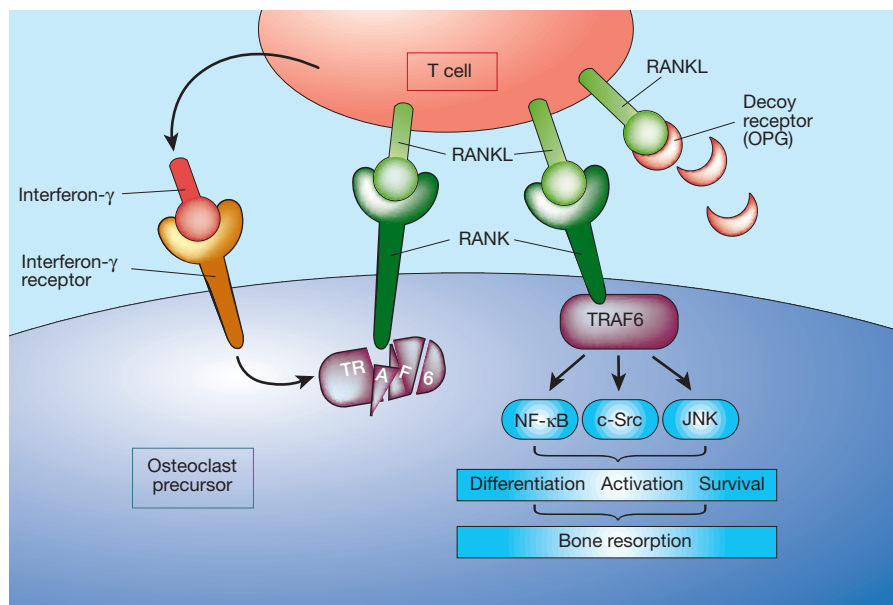


Figure 1 The interface between the immune system and bones. During inflammation, activated T cells express more of the RANKL protein and secrete the interferon- γ molecule. RANKL is important for communication between T cells and dendritic (antigen-presenting) cells. It is also essential for the differentiation, activation and survival of osteoclasts — the cells that break down bone. It binds to RANK on osteoclasts to trigger intracellular signalling pathways that promote osteoclast differentiation and bone resorption. The TRAF6 protein is a key intermediate in these pathways. However, if osteoclasts were activated every time T cells were activated, bone loss would be more common than it is. Takayanagi *et al.*² propose a solution: interferon- γ promotes the degradation of TRAF6, thereby preventing activated T cells from triggering massive bone destruction during inflammation. The effects of RANKL can also be blocked by OPG⁵, a soluble decoy receptor. Together, then, RANKL, interferon- γ and OPG maintain the balance between bone deposition and bone resorption. Note that RANKL is also found on other cells, such as osteoblasts.

formed from bone-marrow-derived precursor cells in response to the coordinated actions of several protein factors. The study of one of these proteins, RANKL (also known as TRANCE, OPGL and ODF), and its binding partners has led to the tales of the immune system and bones becoming more closely intertwined. It seems that the cells of the immune system and the cells that remodel bone regulate each other's function.

RANKL is normally expressed on osteoblasts and activated T cells (Fig. 1). It binds to RANK, which is expressed on osteoclasts and dendritic cells¹ (antigen-presenting cells needed for T-cell-dependent immune responses). When RANK on osteoclasts is activated it sends signals into the cells through 'adapter proteins'. One of these, TRAF6, allows RANKL to activate several signalling pathways (including the transcription factor NF- κ B and the protein kinases JNK and c-Src)⁴. These signalling pathways allow osteoclasts to mature and develop their bone-resorbing functions. A receptor called OPG also binds to RANKL, and acts as a decoy to prevent it from binding to RANK and activating target cells⁵.

So, RANKL is expressed on activated T cells, binds to RANK on osteoclasts, and thereby induces these bone-resorbing cells

to mature and become active. Given all this, one might expect massive bone destruction under most inflammatory conditions — for example, in autoimmune diseases, cancer, allergy, infection and even, perhaps, injury. Indeed, RANKL on activated T cells can activate osteoclasts and trigger bone loss⁶. In arthritic rats, use of the decoy receptor OPG prevents bone destruction (but not inflammation)⁶. Moreover, in mice engineered to lack RANKL, overexpression of this protein on T cells restores some of the normal functions of osteoclasts⁷, suggesting that T cells are important in osteoclast function under non-pathological conditions. But why isn't bone loss more extensive? This is where Takayanagi *et al.*² come in.

These authors have discovered a crucial counter-regulatory mechanism, by which activated T cells can inhibit the RANKL-induced maturation and activation of osteoclasts. Although the T cells involved in inflammatory conditions express RANKL, they also secrete another protein, interferon- γ . Takayanagi *et al.* have found that interferon- γ blocks RANKL-induced osteoclast differentiation *in vitro*. They also show that, in mice engineered to lack the interferon- γ receptor, the induction of inflammatory arthritis results in much more bone destruction than in normal mice. In a nutshell,

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

it seems that interferon- γ — probably amongst other factors — prevents uncontrolled bone loss during inflammatory T-cell responses.

Takayanagi *et al.* suggest that interferon- γ activates the ubiquitin-proteasome pathway — involved in protein degradation — within the osteoclasts, resulting specifically in the degradation of the adapter protein TRAF6 (Fig. 1). They have several reasons for making this proposal. First, osteoclast maturation in the presence of interferon- γ can be 'rescued' by overexpression of TRAF6 in early osteoclast cells. Second, the suppressive effect of interferon- γ on osteoclast formation was markedly decreased in mice lacking proteasome components.

Third, signals downstream of CD40 — a protein that is important in regulating B cells and dendritic cells, and which also signals through TRAF6 — were also attenuated by treatment with interferon- γ . This result also suggests a broader role for interferon- γ . Moreover, TRAF6 is involved in yet more signalling pathways, such as those requiring lipopolysaccharide and interleukin-1 (ref. 8), which are proteins involved in the innate — rather than the adaptive — immune response. It will be interesting to find out whether interferon- γ has effects on these pathways, too.

The fact that there are complex regulatory interactions between bone-remodelling cells and immune cells has broad implications for researchers studying either system. We suggest that the term 'oste-immunology' could be used to describe the interface between these disciplines. It is becoming clear that, without a better understanding of this interface, it will be difficult to prevent or treat many common diseases that affect both bones and the immune system.

Joseph R. Arron is in the Tri-Institutional MD-PhD Program and Yongwon Choi is at the Howard Hughes Medical Institute, Laboratory of Immunology, The Rockefeller University, New York, New York 10021, USA.

e-mails: arronj@rockefeller.edu

choi@rockefeller.edu

- Wong, B. R., Josien, R. & Choi, Y. *J. Leuk. Biol.* **65**, 715–724 (1999).
- Takayanagi, H. *et al. Nature* **408**, 600–605 (2000).
- Rodan, G. A. & Martin, T. J. *Science* **289**, 1508–1514 (2000).
- Wong, B. R. *et al. Mol. Cell* **4**, 1041–1049 (1999).
- Hofbauer, L. C. *et al. J. Bone Mineral Res.* **15**, 2–12 (2000).
- Kong, Y. Y. *et al. Nature* **402**, 304–309 (1999).
- Kim, N., Odgren, P. R., Kim, D. K., Marks, S. C. Jr & Choi, Y. *Proc. Natl Acad. Sci. USA* **97**, 10905–10910 (2000).
- Lomaga, M. A. *et al. Genes Dev.* **13**, 1015–1024 (1999).

Cultural revolution in whale songs

Humpbacks have picked up a catchy tune sung by immigrants from a distant ocean.

The song patterns of humpback whales (*Megaptera novaeangliae*) depend on where they live, with populations inhabiting different ocean basins normally singing quite distinct songs. Here we record a unique and radical song change in the song of humpback whales in the Pacific Ocean off the Australian east coast. Their song was replaced rapidly and completely by the song of the Australian west coast population from the Indian Ocean, apparently as a result of the introduction of only a small number of 'foreign' singers. Such a revolutionary change is unprecedented in animal cultural vocal traditions and suggests that novelty may stimulate change in humpback whale songs.

Male humpback whales sing while migrating to and from their breeding grounds, and when they are at the grounds themselves^{1–3}. Song is thought to be a form of sexual display, but it is not known whether its main purpose is to repel other males or to attract females^{2,4}. All males in a population produce the same song, which changes through time^{2,3,5}, and all singers maintain the changes, implying that there is a cultural transmission and evolution as in some bird songs⁶. Songs across an ocean basin are broadly similar — differences increase with distance — but populations in different oceans separated by continents have apparently unrelated songs^{3,7}.

In the austral winter and spring, humpback whales are found along the east coast of Australia^{3,8}, calving and mating in the waters of the Great Barrier Reef. Using hydrophones suspended from radio-linked buoys and small boats, we recorded humpback whale songs off southeast Queensland during northward and southward migrations between 1995 and 1998, from which 1,057 hours of song were analysed. In 1995 and 1996, the song pattern changed slightly in an evolutionary fashion. But two singers out of 82 were singing a new, completely different song (Fig. 1). In 1997, the new song became more common. Most of the 112 singers produced either the old or the new song, but three used an intermediate song containing themes from both types. By the end of the 1997 southward migration, almost all males had switched songs, and in 1998 only the new song was heard.

The new song was nearly identical to the song of humpback whales migrating along the west coast of Australia in 1996. West and east coast songs are usually very different³ and there is only a small amount of interchange between these populations^{8,9}. The very low incidence of the new song in

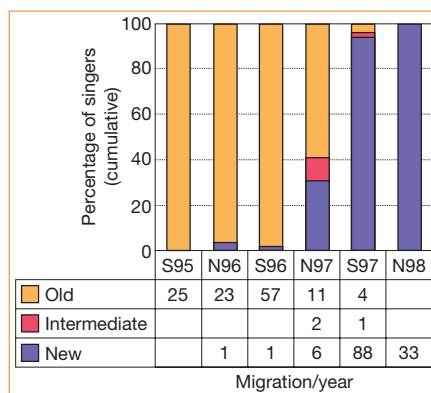


Figure 1 Relative prevalence of different song types in male humpback whales migrating along the east coast of Australia in 1995–98. S, southward; N, northward. The table shows the number of males singing each song type during each migration; the graph shows the relative prevalence of each song type. Most changes in the song patterns occur during the breeding season, when song is being used, and not during the rest of the year when the whales are feeding in high-latitude waters. Over 1,000 hours of song were analysed aurally and spectrographically to determine the song pattern for 252 song bouts.

1996 and the fact that the songs of the two populations evolved independently after 1996 is consistent with the new song being introduced by movement of a small number of singers between populations in 1996.

Humpback whale song shows similarities to song in some birds, particularly the Panamanian yellow-rumped cacique (*Cacicus cela vitellinus*)¹⁰ and village indigo bird (*Vidua chalybeata*)¹¹, in which song repertoires are colony-specific and all individuals have similar repertoires which change with time. But there are no examples of radical song replacement initiated by a small number of immigrant individuals in these or any other species of songbird.

Alloys

A stable binary quasicrystal

All stable quasicrystals known so far are composed of at least three metallic elements^{1–4}. Sixteen years after the discovery of the quasicrystal⁵, we describe a stable binary quasicrystalline alloy in a cadmium–ytterbium (Cd–Yb) system. The structure of this alloy represents a new class of packing of 66-atom icosahedral clusters whose internal structure breaks the icosahedral symmetry. The binary quasicrystal offers a new opportunity to investigate the relation between thermodynamic stability

The process of change in humpback whale song and bird song has been classified as 'cultural evolution', whereby changes in songs are passed among individuals by learning and accumulate over time⁶. The changes we describe in the song of the humpback whales off the east coast of Australia were cultural in that they were due to the learning of a vocal behavioural pattern and not to a mass influx of immigrants. But the rapid and complete replacement of a complex song over a period of less than two years is revolutionary rather than evolutionary, and suggests that novelty drives changes in humpback whale song. To our knowledge, such 'cultural revolution' is unknown in the vocal cultural tradition of any other animal.

Michael J. Noad*, **Douglas H. Cato*†**,
M. M. Bryden*, **Micheline-N. Jenner‡**,
K. Curt S. Jenner‡

*Department of Veterinary Anatomy and Pathology, University of Sydney, New South Wales 2006, Australia

e-mail: mnoad@mail.usyd.edu.au

†Defence Science and Technology Organisation, Pyrmont, New South Wales 2009, Australia

‡Centre for Whale Research, PO Box 1622, Fremantle, Western Australia 6959, Australia

1. Payne, R. S. & McVay, S. *Science* **173**, 585–597 (1971).
2. Winn, H. E. & Winn, L. K. *Mar. Biol.* **47**, 97–114 (1978).
3. Cato, D. H. *Mem. Queensland Mus.* **30**, 277–290 (1991).
4. Tyack, P. *Behav. Ecol. Sociobiol.* **8**, 105–116 (1981).
5. Payne, K., Tyack, P. & Payne, R. in *Communication and Behavior of Whales* (ed. Payne, R.) 9–57 (Westview, Boulder, CO, 1983).
6. Munding, P. C. *Ethol. Sociobiol.* **1**, 183–223 (1980).
7. Winn, H. E. *et al. Behav. Ecol. Sociobiol.* **8**, 41–46 (1981).
8. Chittleborough, R. G. *Aust. J. Mar. Freshwat. Res.* **16**, 33–128 (1965).
9. Baker, C. S. *et al. Mar. Mamm. Sci.* **14**, 721–737 (1998).
10. Trainer, J. M. *Ethology* **80**, 190–204 (1989).
11. Payne, R. B. Z. *Tierpsychol.* **70**, 1–44 (1985).

and quasiperiodic structure, as well as providing a basis for the construction of crystallographic models.

We have previously^{6,7} prepared quasicrystals in alloys of the Cd–Mg–RE type (where RE is a rare-earth metal) by annealing. We systematically eliminated the magnesium component and tested the effect on the phase diagram of the Cd–Yb system. In doing so, we discovered an unknown phase with a congruent melting point of 636 °C at a composition of Cd_{5.7}Yb (where Yb was 15 atom%)⁸. We prepared this alloy of Cd_{5.7}Yb by using an induction furnace and solidified it in an aluminium oxide crucible under an argon atmosphere.

Single crystals of Cd_{5.7}Yb have an

orthogonal unit cell with very large edges, but they did not correspond to any known structure⁹. The powder X-ray diffraction pattern obtained from this alloy identified it as a quasicrystal, and selected-area electron-diffraction patterns verified it had a primitive icosahedral lattice. The quasicrystalline phase was evident in the solidified as well as in the fully annealed state, indicating that the quasicrystalline phase is thermodynamically stable. To our knowledge, this is the first evidence of a thermodynamically stable quasicrystal in a binary system. Unlike all other stable quasicrystals, which show incongruent melting, this new stable quasicrystal has a congruent melting point, as can be seen from the phase diagram^{8,9}.

Figure 1a shows transmission Laue X-ray diffraction patterns taken with incidence along 5-fold, 3-fold and 2-fold axes from a single quasicrystal of size 0.2 mm extracted from the as-cast alloy. The patterns show many sharp reflections with 5-fold, 3-fold and 2-fold symmetry, indicative of a highly ordered quasiperiodicity with icosahedral symmetry persisting over a long range.

The structures of Al–Mn and Al–Zn–Mg quasicrystals can be deduced from their crystalline approximant structures α -Al–Mn–Si and (Al,Zn)₄₉Mg₃₂, respectively^{10,11}. (A crystalline approximant is a phase whose composition is close to or identical to that of a quasicrystal and whose unit cell has atomic decorations that look like a quasicrystal, such as icosahedral clusters of atoms, but which is nonetheless a crystal.) Similarly, the structure of our new quasicrystal can be understood by reference to the compound

Cd₆Yb, which lies just next to the quasicrystal in the phase diagram^{8,9}. Cd₆Yb, which has a cubic phase with space group $Im\bar{3}$ and lattice parameter $a = 1.564$ nm, consists of 168 atoms; as it has only a slightly different composition from the stable quasicrystal (about 0.6 atom%), it serves as a convincing crystalline approximant to the quasicrystal⁹.

The cubic Cd₆Yb structure consists of a body-centred-cubic packing of clusters that have atomic shells with icosahedral symmetry around their centre. Figure 1b shows the four successive shells that create this icosahedral cluster. There are only two types of icosahedral cluster known in quasicrystal structure: the Mackay icosahedral cluster for the Al–Mn class of quasicrystal, and the rhombic triacontahedral cluster for the Al–Zn–Mg class of quasicrystal. Both types contain three different atomic shells, each with an icosahedral symmetry. However, in the case of Cd_{5.7}Yb, four atoms at the centre of the cluster break this icosahedral symmetry. The present quasicrystal can thus be classified as a second type of quasicrystal according to the cluster classification.

Our stable binary quasicrystal will be valuable for studying stabilization mechanisms and thermodynamics in binary phase diagrams. The breakthrough from stable ternary quasicrystals to a binary one will simplify the analysis of experimental results by reducing the parameters for theoretical calculation. It is surprising, given that the crystalline approximant of Cd₆RE forms in most Cd–RE alloys, that a binary stable quasicrystal forms only in a Cd–Yb system. How an internal symmetry-breaking cluster

can build up to form a bulk quasicrystal remains an open question.

A. P. Tsai, J. Q. Guo, E. Abe, H. Takakura, T. J. Sato

National Research Institute for Metals, and CREST, Japan Science and Technology Corporation, Tsukuba 305-0047, Japan

e-mail: aptsai@tamamori.nrim.go.jp

1. Ball, M. D. & Lloyd, D. J. *Scripta Metall.* **19**, 1065–1068 (1985).
2. Tsai, A. P., Inoue, A. & Masumoto, T. *Jpn J. Appl. Phys.* **26**, 1505–1507 (1987).
3. Tsai, A. P., Inoue, A., Yokoyama, Y. & Masumoto, T. *Mater. Trans. Jpn Inst. Metals* **31**, 98–103 (1990).
4. Tsai, A. P. et al. *Phil. Mag. Lett.* **70**, 169–204 (1994).
5. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
6. Guo, J. Q., Abe, E. & Tsai, A. P. *Phil. Mag. Lett.* (in the press).
7. Guo, J. Q., Abe, E. & Tsai, A. P. *Jpn J. Appl. Phys.* **39**, L770–L771 (2000).
8. Massalski, T. B., Okamoto, H., Subramanian, P. R. & Kacprzak, L. (eds) *Binary Alloy Phase Diagrams* 2nd edn, Vol. 2, 996 (ASM International, Gaithersburg, 1990).
9. Palenzona, A. J. *Less-Common Met.* **25**, 367–372 (1971).
10. Elser, V. & Henley, C. L. *Phys. Rev. Lett.* **55**, 2883–2886 (1985).
11. Henley, C. L. & Elser, V. *Phil. Mag. Lett.* **B 53**, L59–L66 (1986).

Development

DNA methylation in *Drosophila melanogaster*

Certain cytosine residues of eukaryotic DNA are methylated in inactive regions of the genome. For a long time the fruitfly *Drosophila* was thought to be an exception^{1–4}, but now the evidence points to the existence of a functional DNA-methylation system in *Drosophila* as well^{5–9}. Here we show that DNA is methylated, but that *Drosophila* genomic methylation is restricted to the early stages of embryonic development.

We analysed genomic DNA from *Drosophila melanogaster* embryos (0–4 h old) by high-performance liquid chromatography (HPLC), which revealed a weak signal for 5-methylcytosine (Fig. 1a), consistent with low levels of DNA methylation.

We then analysed genomic DNA at different stages of development by nearest-neighbour analysis and thin-layer chromatography. We investigated the methylation of CpG dinucleotides at DNA sites cut by the restriction enzyme *FokI* (specific for the sequence GGATGN_{9–13}) and found a weak signal from embryonic DNA (Fig. 1b; Table 1).

Extension of nearest-neighbour analysis to non-CpG methylation revealed a distinct 5-methylcytosine spot in embryonic-DNA preparations (Fig. 1c). Most methylation was seen at CpT, with some at CpA and CpC (Table 1). Genomic DNA methylation was prevalent in young (1–2 h) embryos (Fig. 1c), but less marked in older (15–16 h) embryos (Fig. 1d); only trace amounts of 5-methylcytosine were found in isolated ovaries (Fig. 1e). We confirmed these results by using a different restriction

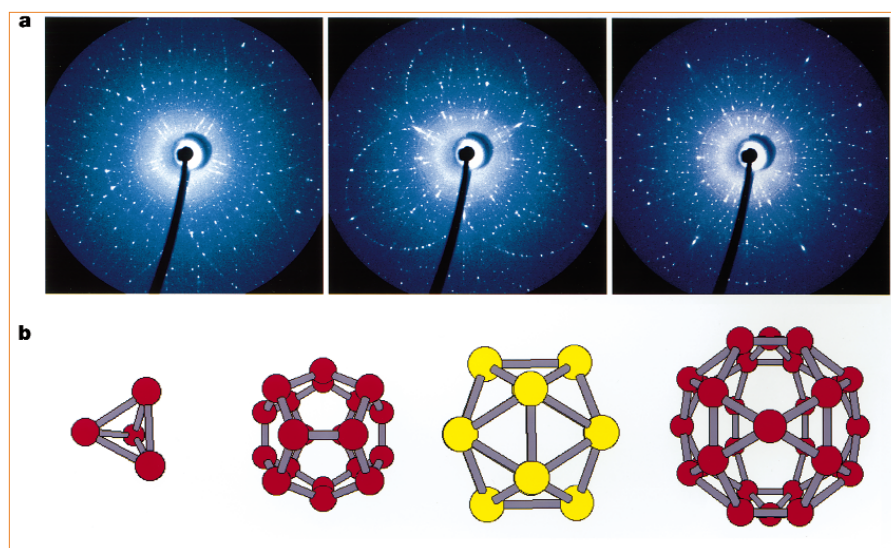


Figure 1 Cd_{5.7}Yb quasicrystal. **a**, Transmission Laue X-ray diffraction patterns along 5-fold, 3-fold and 2-fold axes (respectively, left to right) from a single quasicrystal. Based on its orientation, the quasicrystal was confirmed to have icosahedral symmetry. Patterns were obtained from a single grain, indicating that the quasicrystal structure is stable enough to create a bulk form. **b**, Decoration of the icosahedral cluster in the quasicrystalline Cd_{5.7}Yb alloy, as deduced from the cubic Cd₆Yb crystalline approximant. The first shell is created by four Cd atoms around the cluster centre, the second consists of 20 atoms forming a dodecahedron, the third is an icosahedron of 12 Yb atoms, and the fourth is a Cd icosidodecahedron obtained by placing 30 Cd atoms on the edges of the Yb icosahedron. In all, the icosahedral cluster consists of 66 atoms. Cd, red atoms; Yb, yellow atoms.

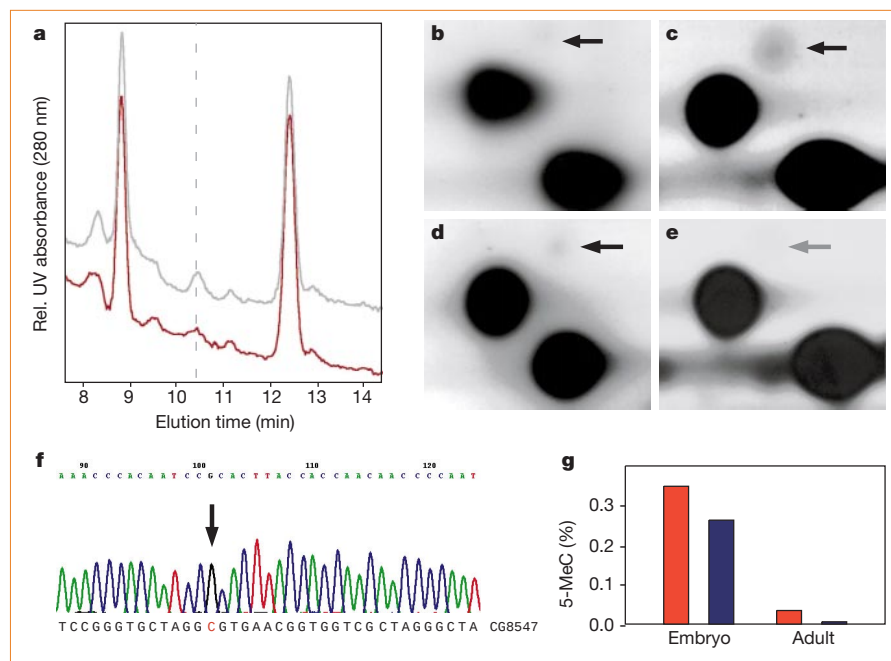


Figure 1 Methylation analysis of *Drosophila* genomic DNA from wild-type (*Oregon R*) flies. Genomic DNA from washed, dechorionated embryos was digested with ribonucleases A and T1 to remove RNA. **a**, Analysis by isocratic reverse-phase HPLC¹⁵. Red trace, DNA from 0–4-h staged embryos; grey trace, same sample spiked with desoxy-5-methylcytosine for calibration (dashed line). **b–e**, Nearest-neighbour analysis¹² of ³²P-labelled samples run on a thin-layer chromatogram after digestion with *FokI*. Arrows, 5-methylcytosine. **b**, CpG, and **c**, CpT methylation in 1–2-h embryos; **d**, CpT methylation in 15–16-h embryos, and **e**, in isolated adult ovaries. **f**, Sequence of genomic¹² bisulphite-deaminated DNA from early embryos. 5-methylcytosine (arrow) cannot be deaminated and so remains the only base sequenced as cytosine (or guanine on the complementary strand). Alignment of the deaminated sequence (top) shows that it is derived from the *Drosophila* CG8547 locus (bottom). 5-Methylcytosine is in the putative promoter region. **g**, Cytosine methylation in early embryos and in adults, as determined by nearest-neighbour analysis (red) and by sequencing of deaminated genomic DNA (blue).

enzyme (*Bst*NI) for nearest-neighbour analysis (Table 1).

We calculated that the average amount of (cytosine-5) methylation was 0.4% in young (1–2 h) embryos and 0.1% in older (15–16 h) embryos. Although low, this methylation is significant because no 5-methylcytosine was detectable after restriction-enzyme digestion of unmethylated controls produced by the polymerase chain reaction.

Isolated ovaries had very little DNA methylation, probably because of a lack of the methyltransferase enzyme. We analysed DNA methylation in male and female adult flies and found 5-methylcytosine levels comparable to those in isolated ovaries.

In some preparations from adult flies, methylation was concentrated at *Bst*NI-cut sites (data not shown), but as *Bst*NI recognizes the bacterial methylation sequence

CCWGG, this was presumably due to external bacterial contamination of the fly preparations. None of the DNAs prepared from internal tissue (dechorionated embryos, isolated ovaries) showed this concentration of 5-methylcytosine at *Bst*NI-cut sites (Table 1).

We confirmed the presence of 5-methylcytosine in the *Drosophila* genome by deaminating genomic DNA from young embryos and from adults with bisulphite and then sequencing random fragments. Under our conditions, 5-methylcytosine is protected against deamination and so remains as the only base to be sequenced as cytosine. Of 12 kilobases (71 independent clones) of deaminated embryonic DNA and 13 kilobases (57 independent clones) of deaminated adult DNA, almost all sequences (over 90%) could be unambiguously aligned to the *Drosophila* genome sequence¹⁰.

We found that eight cytosine residues representing genomic 5-methylcytosine in embryonic DNA and respective fragments could be aligned to the *Drosophila* genome sequence (Fig. 1f). In contrast, no cytosine residues representing genomic 5-methylcytosine were found in adult DNA. The corresponding cytosine-methylation level (0.3% for embryos and 0.0% for adults) is in good agreement with values determined by nearest-neighbour analysis (Fig. 1g).

Previous attempts to identify DNA methylation in the fly have focused mainly on late stages of development^{1–4}. Our results show that DNA methylation in *Drosophila* predominates during early embryonic development and decreases at later stages, presumably as a result of reduced methyltransferase expression⁹. Tiny amounts of 5-methylcytosine have been found in adult *Drosophila* DNA¹¹ but the significance of this was unclear until now. Surprisingly, we found only a small fraction of the 5-methylcytosine in CpG dinucleotides, which may have contributed to previous failures to detect DNA methylation in the fly, as many of these experiments focused on methylation of CpG. Methylation of other dinucleotide residues during the early stages of vertebrate development¹² when Dnmt2 is expressed^{13,14} could also have been a factor — attempts to demonstrate DNA-methyltransferase activity for Dnmt2 might have been compromised by focusing on CpG rather than methylation at other sites.

Frank Lyko*, Bernard H. Ramsahoye†, Rudolf Jaenisch*‡

*Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

e-mail: jaenisch@we.mit.edu

†John Hughes Bennett Laboratory, Department of Oncology, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK

‡Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Table 1 Quantification of DNA methylation

Restriction enzyme	Sample	Dinucleotide	5-MeC (c.p.m.)	Background (c.p.m.)	Cytosine (c.p.m.)	5-MeC (%)
<i>FokI</i>	1–2-h embryos	CpG	160	140	13,757	0.1
		CpT	318	48	36,706	0.7
		CpA	147	58	30,438	0.3
		CpC	175	74	40,972	0.2
<i>FokI</i>	15–16-h embryos	CpG	108	60	59,949	0.1
		CpT	186	66	59,195	0.2
		CpA	129	68	76,795	0.1
		CpC	144	91	195,910	0.0
<i>FokI</i>	Ovary	CpG	234	226	104,473	0.0
		CpT	251	94	137,141	0.1
		CpA	211	130	228,680	0.0
		CpC	120	125	374,100	0.0
<i>Bst</i> NI	1–2-h embryos	CpT	192	49	73,799	0.2
		CpA	165	60	52,143	0.2
<i>Bst</i> NI	15–16-h embryos	CpT	161	57	101,575	0.1
		CpA	68	67	24,725	0.0
<i>Bst</i> NI	Ovary	CpT	142	69	175,071	0.0
		CpA	57	56	29,576	0.0

Genomic DNA was digested by one of the indicated restriction enzymes and (cytosine-5) DNA methylation was analysed in CpN dinucleotides by ³²P-labelling of dNTP. TLC spots corresponding to 5-methylcytosine (5-MeC), cytosine or background were counted. Methylation of 5-MeC was calculated as a percentage of total counts incorporated into cytosine and 5-MeC after adjusting for background counts. As negative controls, we did a similar analysis on a PCR product containing *FokI* and *Bst*NI sites, including a mock ribonuclease digestion step to control for the presence of exogenous methylated DNA in enzyme preparations and buffers. No 5-MeC was detected in any of these controls.

1. Rae, P. M. & Steele, R. E. *Nucleic Acids Res.* **6**, 2987–2995 (1979).
2. Bird, A. P. & Taggart, M. H. *Nucleic Acids Res.* **8**, 1485–1497 (1980).
3. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. *FEBS Lett.* **146**, 148–152 (1982).
4. Patel, C. V. & Gopinathan, K. P. *Anal. Biochem.* **164**, 164–169 (1987).
5. Hung, M. S. *et al. Proc. Natl Acad. Sci. USA* **96**, 11940–11945 (1999).
6. Tweedie, S. *et al. Nature Genet.* **23**, 389–390 (1999).
7. Wade, P. A. *et al. Nature Genet.* **23**, 62–66 (1999).
8. Lyko, F. *et al. Nature Genet.* **23**, 363–366 (1999).
9. Lyko, F., Whittaker, A. J., Orr-Weaver, T. L. & Jaenisch, R. *Mech. Dev.* **95**, 215–217 (2000).
10. Adams, M. D. *et al. Science* **287**, 2185–2195 (2000).
11. Achwal, C. W., Ganguly, P. & Chandra, H. S. *EMBO J.* **3**, 263–266 (1984).
12. Ramsahoye, B. H. *et al. Proc. Natl Acad. Sci. USA* **97**, 5237–5242 (2000).
13. Yoder, J. A. & Bestor, T. H. *Hum. Mol. Genet.* **7**, 279–284 (1998).
14. Okano, M., Xie, S. & Li, E. *Nucleic Acids Res.* **26**, 2536–2540 (1998).
15. Ramsahoye, B. H., Burnett, A. K. & Taylor, C. *Blood* **87**, 2065–2070 (1996).

Analytical dynamics

Numismatic gyrations

The familiar shuddering motions of spinning coins as they come to rest are not at all intuitive. Moffatt's analysis (*Nature* **404**, 833–834; 2000) identifies air viscosity as the causative factor in coin jitter, so we tested this hypothesis by studying coins spinning in a vacuum. We discovered that the presence of air has little effect on the final motions of the coins, indicating that slippage and friction between the coin's edge and the supporting surface might cause the vibrations that accompany the end of the spin.

Casual observation of various objects spun on a tabletop indicates that compression of trapped air does not qualitatively affect the complex motions of spinning disks. We noted that a ring-shaped bell-jar lid, a short cylinder or the lid of a shoe-polish can — tested with either the rim or the flat side down — show a comparable behaviour: they spin on edge, topple over, then wobble to a shuddering halt. The universality of this motion is surprising in light of the air-viscosity mechanism proposed by Moffatt. As rings do not trap air the way solid disks do, these objects should generate shear forces of different magnitudes. The similar kinetic behaviour of these objects appears to contradict a decisive role for air viscosity.

The Dutch 2.5-guilder coin has magnetic properties that allow it to be spun with a precise frequency on a magnetic stirrer. We placed the coin in a glass desiccator that had a slightly concave bottom, brought it to a spin of approximately 10 Hz, and observed the motions of the slowing coin after the desiccator was lifted carefully from the stirring platform. The desiccator could be

evacuated to less than 1 mtorr of air pressure.

Coins *in vacuo* spun on average for 12.5 s; coins in air spun on average for 10.5 s (average of 10 observations each). This difference in time can be attributed to a difference in the time the coin was spinning upright on its edge. The time from the onset of tumbling to standstill did not differ markedly and was about 4 s under both conditions. With or without air, the coin displayed the same characteristic final motions. We conclude that the presence or absence of air may have some effect on the upright duration of the spin, but has little effect on the final whirling motions that bring coins to rest. In contrast, Moffatt's analysis would predict a very long wobbling time for a coin in a vacuum.

We propose an alternative explanation for the jerking motions with which coins lose their spin. A coin toppling from rotation on edge preserves its rotational energy so that the axis of rotation changes from the plane of the coin to one perpendicular to the coin. The coin now must wobble on its edge. As Moffatt indicates, the friction is minimal when the point of contact between the supporting surface and the wobbling coin describes a circle with radius $R\cos(\alpha)$ (see his Fig. 1). But the coin is not free to choose any rotation speed. The gravitational force supplies a moment that interacts with the spin moment and the wobble moment. As a result, the coin is subject to precessing forces that rub the coin's edge in a jerking motion against the tabletop. We believe that this sliding friction temporarily lifts the coin, moving the point of contact between edge and supporting surface in a rapid staccato. It is this friction that brings the coin to a final rest.

The role of surface friction can be readily confirmed with the toy that inspired Moffatt's analysis. When placed on a table rather than on its slippery platform, Euler's disk rapidly comes to rest, illustrating the influence of the roughness of the supporting surface on the spinning time. Air viscosity may play a role in stopping 'theoretical' coins. Real-world coins, thrown on a table, do not need a finite-time singularity to control their spin. Edges rubbing against the tabletop explain the rapid dissipation of monetary momentum.

Ger van den Engh*, Peter Nelson†, Jared Roach‡

*Institute for Systems Biology, Seattle, Washington 98105, USA

e-mail: engh@biotech.washington.edu

†Department of Bioengineering, University of Washington, Seattle, Washington 98195, USA

‡University of Utah, Salt Lake City, Utah 84132, USA

Moffatt replies — It is true that there are a number of possible dissipative mechanisms

for the rolling disk in addition to viscous dissipation in the surrounding air: vibration of the supporting surface, rolling friction due to plastic deformation at the point of rolling contact, and, as suggested by van den Engh *et al.*, dissipation due to slipping rather than rolling. The 'adiabatic' equation that I used, relating the precessional angular velocity Ω to the angle α , is valid only under the rolling condition, and experiments indicate that this condition is indeed satisfied for the 'toy' Euler's disk rolling on a flat, smooth horizontal glass plate placed on a firm table (V. A. Vladimirov, personal communication). I believe therefore that slipping does not occur in this case.

The problem really is to identify the dominant dissipative mechanism, for a given disk and a given surface, and then to evaluate the associated rate of dissipation of energy as a function of the angle α (which is proportional to the energy). If this rate of dissipation of energy turns out to be proportional to a power of α , where the exponent of this power, λ say, is less than one, then, under the adiabatic approximation, a finite-time singularity (for which Ω becomes infinite) will occur.

The air-viscosity mechanism I described yields $\lambda = -2$ (note that air viscosity is relatively insensitive to pressure, so that partial evacuation of the vessel in which the disk experiment is conducted should have only a small effect). An improved theory that takes account of oscillatory Stokes layers on the disk and supporting surface (L. Bildsten, personal communication) yields $\lambda = -5/4$. If 'rolling' friction is assumed to dissipate energy at a rate proportional to Ω , then $\lambda = -1/2$. Careful experiments under a variety of conditions should distinguish between these various possibilities.

I chose to focus on viscous dissipation because that is the only mechanism for which a fundamental (rather than empirical) description is available, namely that based on the Navier–Stokes equations of fluid dynamics. The fact that the air-viscosity mechanism exhibits the strongest singularity as α tends to zero suggests that this mechanism will always dominate when α is sufficiently small. For larger α and smaller disks (such as the 2.5-guilder coin), rolling friction is an equally plausible candidate (A. Ruina, personal communication), but determination of the associated rate of dissipation of energy (in terms of the physical properties of the disk and the surface) involves solution of the equations of (possibly plastic) deformation in both solids at the moving point of rolling contact, a difficult problem, which, so far as I am aware, still awaits definitive analysis.

H. K. Moffatt

Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, Cambridge CB3 0EH, UK
e-mail: hkm2@newton.cam.ac.uk

Electronics using hybrid-molecular and mono-molecular devices

C. Joachim*, J. K. Gimzewski† & A. Aviram‡

* Centre d'Elaboration de Matériaux et d'Etudes Structurales—Centre National de la Recherche Scientifique, 29 rue J. Marvig, 31055 Toulouse Cedex, France

† IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

‡ IBM Research, T. J. Watson Research Center, Route 134, Yorktown Heights, New York 10598, USA

The semiconductor industry has seen a remarkable miniaturization trend, driven by many scientific and technological innovations. But if this trend is to continue, and provide ever faster and cheaper computers, the size of microelectronic circuit components will soon need to reach the scale of atoms or molecules—a goal that will require conceptually new device structures. The idea that a few molecules, or even a single molecule, could be embedded between electrodes and perform the basic functions of digital electronics—rectification, amplification and storage—was first put forward in the mid-1970s. The concept is now realized for individual components, but the economic fabrication of complete circuits at the molecular level remains challenging because of the difficulty of connecting molecules to one another. A possible solution to this problem is ‘mono-molecular’ electronics, in which a single molecule will integrate the elementary functions and interconnections required for computation.

Aviram and Ratner¹ suggested that a single molecule with a donor–spacer–acceptor (d–s–a) structure (see 1 in Fig. 1c) would behave as a diode when placed between two electrodes: electrons can easily flow from the cathode to the acceptor, and electrons from the donor are then transferred to the anode. The working principle of this device is analogous to that underlying the “valve” effect introduced by Shockley 60 years ago², but involves manipulating the electronic wavefunction of the metallic electrodes extending through the d–s–a molecule, rather than the carrier density in a semiconductor material. Such hybrid molecular electronic (HME) devices, comprising molecules embedded between several electrodes, thus differ radically from bulk-material-based molecular electronic technologies found in applications such as dye lasers, light-emitting diodes, liquid-crystal displays, and soft plastic transistors. However, the design of functional devices and machines based on the molecular electronics concept poses the challenge of integrating the functions required for advanced processing, particularly computing, within the same molecule in a mono-molecular electronics (MME) approach.

Owing to the lack of suitable technologies for establishing electrical contact between individual molecules, experimental investigations of the fundamental processes involved in electron transfer through molecules have long focused on gas-phase and liquid-phase systems. Although these approaches cannot result in working nanoscale devices³, they have established that a simple shift in the energy of a molecular orbital—induced, for example, by a photoisomerization—can fine-tune the intramolecular electron transfer rate through an individual molecular bond connecting two redox centres. Experimental studies involving averaging over many isolated molecules have, for example, demonstrated long-range through-bond electron transfer processes in molecule 2 (Fig. 1a)^{4,5}, intramolecular light-induced conformation changes in molecule 3 (Fig. 1b)⁶, and intramolecular interference effects in molecule 4 (Fig. 1b)⁷.

Electrical addressing of molecules

Langmuir–Blodgett (LB) and self-assembly fabrication techniques⁸ can be used on many molecules to form organized molecular monolayers on suitable substrates. This allowed Mann and Kuhn to investigate long-range tunnelling through alkane chains in ordered LB monolayers⁹. Similarly, sandwiching molecule 5

(Fig. 1c) between differing metallic electrodes allowed the first (albeit somewhat ambiguous) observation of rectification effects in a molecular monolayer¹⁰, confirmed by subsequent work¹¹ using molecule 6 (Fig. 1c) sandwiched between electrodes made from the same material.

The scanning tunnelling microscope (STM) enables controlled two-terminal measurements, and its development has thus allowed new experimental approaches for demonstrating and probing electron transport through individual molecules¹². Examples include the electrical single-atom switch realized using a Xe atom¹³ at cryogenic temperatures and the first experimental determination of the electrical contact point of a single C₆₀ molecule (ref. 14). The resistance of $R = 55 \text{ M}\Omega$ obtained in the C₆₀ experiment corresponds to an electronic transparency (ease of transmission) of $T = 2.3 \times 10^{-4}$, with T being approximately proportional to the square of the inter-electrode electronic coupling introduced by the molecule compared to the corresponding vacuum gap. Ohmic dissipation in the electrodes is one way to evaluate T from the macroscopically measurable quantity R , as $T = h/2e^2 R^{-1}$. However, owing to R being the resistance of the metal–C₆₀–metal tunnel junction, rather than that of the C₆₀ molecule, it cannot be used to define molecular conductivity.

STM measurements on C₆₀ have revealed linear current–voltage (I – V) characteristics at low applied bias voltages¹⁴, indicating that the system behaves similarly to a metal–vacuum–metal tunnel junction at low bias voltage. However, ‘squeezing’ a C₆₀ molecule by applying a small force in the nanonewton range with a metallic STM tip results in a shift of the molecular orbital levels¹⁴. This mechanically induced shift provides a means to modulate tunnelling through a single C₆₀ molecule, even in a nonresonant tunnelling regime, and to change its resistance up to a limiting value approaching the quantum of resistance, $h/2e^2$ ($\sim 12.9 \text{ k}\Omega$) (ref. 14). The phenomenon, mechanically modulated virtual-resonant tunnelling (VRT), has also been used to design an electromechanical amplifier (Fig. 2a), by making use of the ability to mechanically reduce molecular level degeneracy¹⁵. Specific to highly degenerate systems, the electromechanical amplifier is the tunnelling equivalent, on the nanometre scale, of the Shockley “valve” effect².

A variety of other techniques, such as experiments based on Coulomb blockade¹⁶, nanopore¹⁷, break junction^{18,19}, electro-deposition²⁰ and nanolithography^{21,22}, have been used to determine

the electronic transparency of single molecules whose configuration approaches the planar wiring configuration encountered in semiconductor devices. Coulomb blockade work on molecule **7** (Fig. 1a), for example, used a tunnel-capacitor model for the metal–molecule–cluster STM junction used to estimate a tunnelling resistance R from the experimental data and extract the corresponding transparency T , yielding $R = 9 \text{ M}\Omega$ and $T = 1.4 \times 10^{-3}$ (ref. 16). The ‘nanopore’ technique consists of fabricating holes of very small diameter in a suspended silicon nitride membrane equipped with a

metallic electrode in its base. The pores are then filled with molecules and a second metallic electrode is evaporated on top of the device¹⁷. For molecule **8**, Fig. 1a, nonlinear two-terminal I – V characteristics were recorded, and $R = 150 \text{ M}\Omega$ was estimated at low voltage¹⁷. Break junctions involve the gentle fracture of a micro-fabricated electrode in its centre by mechanical deformation while measuring the resistance of the metallic wire junction²³. Its application to single molecules is difficult because a liquid evaporation step is required after formation of the junction, and the

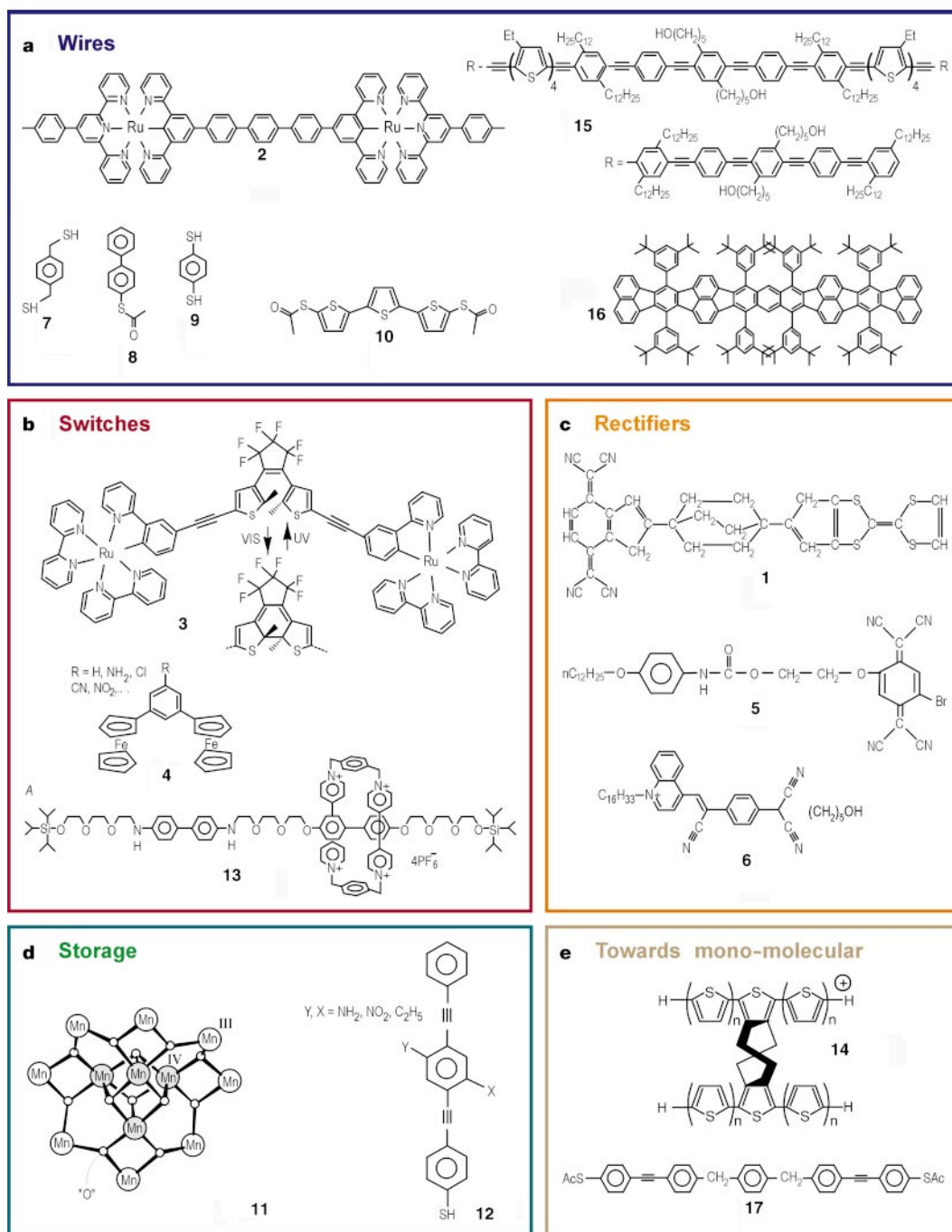


Figure 1 The molecules described in the text. **a**, Wires; **b**, hybrid molecular electronic (HME) switches; **c**, HME rectifiers; **d**, storage; and **e**, two molecules that show promise for mono-molecular electronics. Only the manganese acetate molecule **11** of formula $(\text{Mn}_{12}\text{O}_{12}(\text{CH}_3\text{COO})_{16}(\text{H}_2\text{O})_4)$ is not completely represented, for clarity⁴⁸. All these molecules have been synthesized apart from the historical design **1** of a d–s–a molecular rectifier¹. Molecules **14** and **15** have not yet been characterized. **14** is the first proposal of an intramolecular transistor and **17** is an intramolecular quantum well.

conformation and the exact number of interconnected molecules remain essentially inaccessible¹⁹. Nevertheless, measurements have provided estimates of $R = 22 \text{ M}\Omega$ ($T = 5.9 \times 10^{-4}$) for a junction containing molecule **9** shown in Fig. 1a (ref. 18). In the case of molecule **10** (Fig. 1a) and its dimer, values of $R = 12.5 \text{ M}\Omega$ ($T = 1.03 \times 10^{-3}$) and $R = 160 \text{ M}\Omega$ ($T = 7.7 \times 10^{-5}$) were obtained, respectively¹⁹.

Electro-deposition techniques²⁰ use suspended electrodes in a pseudo-planar configuration to trap molecules electrostatically in and onto the junction. Interrupting the trapping field at the first observed increase of the current in the circuit leads to the deposition of a finite number of molecules in the junction. This approach is similar to break junctions, but there are no imaging methods available to verify the actual number and orientation of the molecules in the junction. Using this configuration, an oligonucleotide straddling an 8-nm nanojunction was found to be negligibly electrically conducting up to an applied potential of several volts²⁴.

Nanolithography has made important contributions towards using mesoscopic electrodes^{25,26} and nanojunctions for HME²⁷. The measurement of the conductance of single-wall carbon nanotubes (SWCNTs) represented the observation of molecular-scale objects electrically interconnected in a full planar configuration, together with images of the position and conformation of the macromolecule^{25,26}. SWCNTs with appropriate helicity were found to transport electrons ballistically²⁵, that is, their electronic transparency is $T = 1$. Consequently, the conductance of the metal–SWCNT–metal junction is determined by the metal–SWCNT contacts. It has generally been difficult to decrease contact resistances in these systems below $100 \text{ k}\Omega$ (ref. 28), but the use of catalysts to pattern and grow SWCNTs directly onto electrodes has recently resulted in two-terminal contact resistances as low as a few kilo-ohms (ref. 29).

At low voltages, two regimes of elastic transport explain the large scatter of transparency values measured on single molecules. (Transport can involve either electrons or quasi-particles, such as polaron and soliton carriers.) For $T > 1$, the transport regime is ballistic: the molecular levels are in resonance with the Fermi level of the electrodes, and there is complete electronic delocalization along

the molecule. In the case of SWCNTs at low temperatures³⁰, the transport coherence length can extend to as much as 500 nm (Fig. 3). Values of $T < 1$ indicate that either the molecule is differently bound to the electrodes or that the Fermi level is located within the gap between the highest occupied (HOMO) and the lowest unoccupied (LUMO) molecular orbitals. In the first case, the tunnelling regime is determined by the metal–molecule contact, which can be optimized by controlling the surface chemistry. To overcome the second situation, the molecular orbitals need to be electronically coupled to the electrodes to stabilize a tunnel pathway for the carriers that is more efficient than vacuum. This pathway builds up from the constructive or destructive superposition of tunnel channels. For C_{60} adsorbed on the $\text{Au}(110)$ surface, this involves 36 of a total of 240 molecular orbitals¹⁴. For systems with $T < 1$ and involving a molecule symmetrically chemisorbed on the electrodes, such as C_{60} bound to two similar electrodes, two-terminal measurements exhibit linear I – V characteristics when V is much lower than the molecule's effective barrier height^{14,19}. This Simmons tunnelling regime is analogous to electron tunnelling between metals at low voltages, also resulting in linear I – V curves. Nonlinear I – V curves usually occur at higher voltages in systems with $T > 1$ involving a physisorbed molecule^{31,32}, owing to the contact tunnel barriers at the electrodes dominating the behaviour of the system²⁵.

Calculations support a very fine dependency of the metal–molecule–metal junction resistance on the chemical structure of the molecule^{33,34}, its conformation in the junction³⁵, and its chemical binding to the electrodes^{33,36,37}. In these calculations, the electronic and mechanical description of the junction includes the atomic structure of the electrode and the molecular conformation obtained from semi-empirical³⁸ or, more recently, from density-functional approaches³⁹. The tunnelling current intensity is usually derived from the Landauer formula¹⁴, with the scattering problem being solved using a spatial propagator⁴⁰ or a Lippman–Schwinger technique through its kernel (Green's function)⁴¹. Alternatively, the current intensity can be calculated from a time-dependent approach that relates the electron transfer process through the

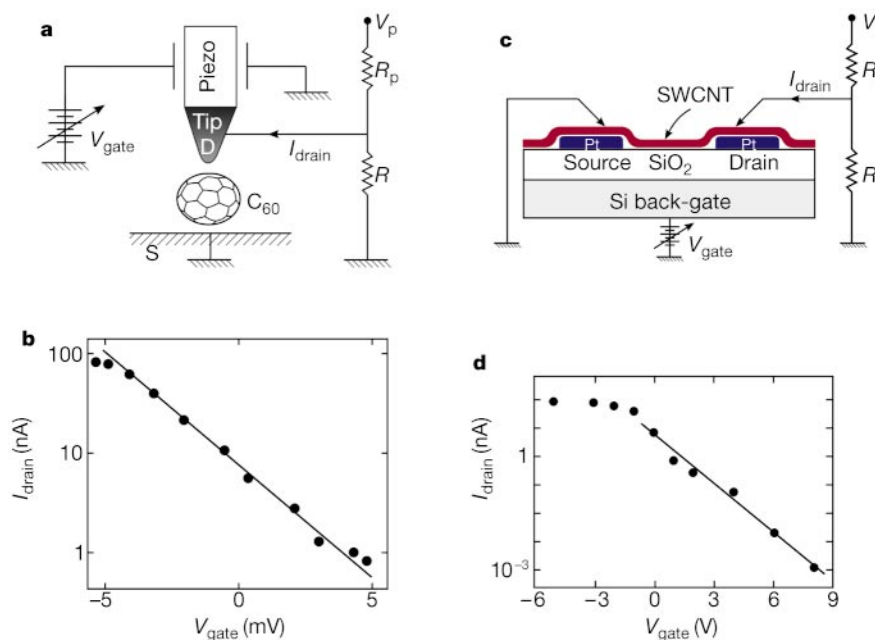


Figure 2 The first two active three-terminal devices in molecular electronics. **a, c**, Electrical circuits of **a**, the electromechanical single C_{60} amplifier (in its original scanning tunnelling microscope version¹⁹) and **c**, the single-wall carbon nanotube (SWCNT) channel transistor⁶³ with their respective $I_{\text{drain}} = f(V_{\text{gate}})$ characteristics (S, source; D, drain; I_{drain} , the tunnel current intensity through the molecule). From **b**, the transconductance $dI_{\text{drain}}/dV_{\text{gate}}$ of the C_{60} amplifier is $g = 3.9 \times 10^{-6} \text{ A V}^{-1}$ and from **d**, that of the SWCNT transistor is $g = 8 \times 10^{-9} \text{ A V}^{-1}$. Those values must be compared to $g = 30 \times 10^{-3} \text{ A V}^{-1}$ for a bipolar transistor and $g = 3 \times 10^{-3} \text{ A V}^{-1}$ for a vacuum-tube triode. The channel of these HME devices (**a, b**) is truly molecular in scale, but not the grid.

molecule to the rate at which the junction electrodes deliver tunnelling electrons to the molecule⁴².

Single hybrid molecular devices

Quasi-ideal rectification characteristics in nanometre-scale objects have been experimentally observed, by sliding an STM tip along a SWCNT⁴³ and in electrically contacted MWCNTs⁴⁴. More recently, a single, planar SWCNT comprising tube sections with different tube helicities, one conducting and the other semiconducting and with a kink at the intervening junction, was shown to exhibit rectification behaviour on the molecular scale⁴⁵. Molecular oligomers should also behave as d-s-a HME rectifiers, with dimensions well below 10 nm (ref. 1), and experimental proof for the validity of the d-s-a design principle has been obtained^{10,11}. But despite these promising advances, many challenges remain: current technologies do not yet permit the realization of planar single-molecule d-s-a rectifiers fabricated with nanometre-scale dimensions, and whether those SWCNT devices can be scaled down by reducing the inter-electrode distance well below 10 nm remains an open question⁴⁶. Similarly, rectification effects observed in nanopores¹⁷ using molecule **8** (Fig. 1a) and in organic heterostructures⁴⁷ still involve several thousand molecules, making it difficult to assess to what extent scaling to a single-molecule level will be possible.

Hysteresis behaviour is a useful property for information storage applications. Manganese acetate (molecule **11** in Fig. 1d), with a total spin magnetic moment of 20 μ_B is the first molecule shown to exhibit a hysteresis cycle under cryogenic conditions⁴⁸. Because magnetic hysteresis is usually seen as a collective property of bulk material, magnetic phenomena associated with single atoms⁴⁹ or molecules⁵⁰ have attracted much interest in the context of device scaling and quantum micro-reversibility⁵¹. An alternative approach to storage is to build devices that display negative differential resistance (NDR)⁵², that is, a negative slope in their I - V curves like the one exhibited by a tunnel diode. Introduced in a two-terminal device, molecules exhibiting NDR play the equivalent role of a nonlinear material in photonic Fabry-Perot resonators showing a hysteresis cycle⁵³. NDR with a high peak-to-valley ratio was observed⁵⁴ for molecule **12** (Fig. 1d).

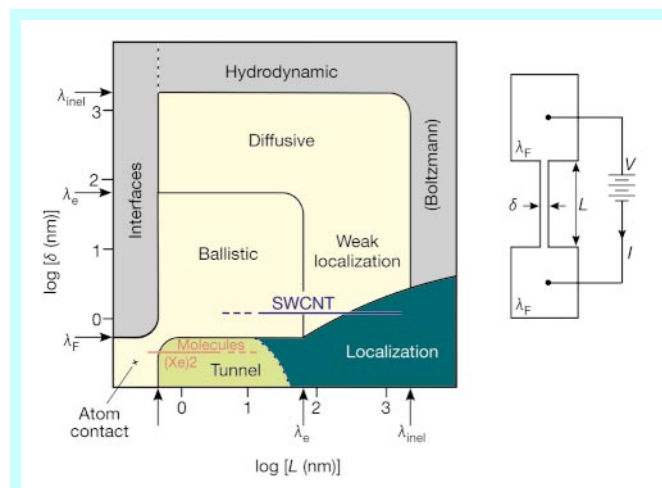


Figure 3 Regimes of electronic transport as a function of the wire width δ and length L . λ_F is the de Broglie carrier wavelength in the contact electrodes (away from the constriction), λ_e the elastic mean free path in the wire and λ_{inel} the inelastic mean free path in the wire. Characteristic orders of magnitude for λ_F , λ_e and λ_{inel} are taken for noble metals at low temperature. Hydrodynamic (Boltzmann), diffusive (weak localization) and ballistic regimes have been well studied in the past for metals and semiconductors. Ballistic and weak localization regimes are now being studied on SWNTs²⁵, atomic metallic wires²³ and the tunnel regime on single molecules^{14,18,19,76} and short atomic wires⁹⁶.

Suggestions^{55–58} for single-molecule molecular switches have been put forward for some time, but relatively few have been synthesized^{6,59}. In an electronic circuit, switching is based on an intrinsic molecular property involving a bistable change of internal structure (such as a conformation change⁵⁷ or a unimolecular reaction^{6,59}), which induces a modification of T . For example, the photochromic switching of molecule **3** (Fig. 1b) is possible because a photo-induced molecular orbital “up-shift” in the “on” state⁶ favours efficient electron transfer through **3**, making this phenomenon an intramolecular analogue of the solid state “valve” effect. Rotaxanes **13** (Fig. 1b) have also been proposed as mechanical molecular switches⁶⁰; the macromolecular ring can move along the molecule’s central ‘axle’ and occupy one of two or more metastable positions or ‘stations’ along the axle, where one position provides a high- T and the other a low- T state. The gating used for these two switches has a macroscopic origin (light for the photochromic effect, pH and light control for the rotaxane), but the development of near-field integrated optics⁶¹ may provide a more controlled means to optically gate molecular switches on the mesoscopic scale.

Proposals of HME transistors are numerous⁶², yet their realization⁵⁵ remains challenging because logic circuit applications require a three-electrode configuration with high gain. SWCNTs have been used as the channel of a field-effect transistor⁶³, but the device is still mesoscopic in dimension because it operates by controlling the Schottky barrier height at the single-wall nanotube (SWNT)–electrode contacts (Fig. 2b). Plans to move the control of transparency T from the contact between the active molecule and the electrodes to the active molecule itself involve the electromechanical amplification effect of C_{60} (Fig. 2a), which might be planarized⁶⁴ using nano-electromechanics (NEMS)⁶⁵. Oxidation–reduction processes provide another way of controlling the transparency of a molecular tunnel pathway. For instance, conjugated molecules such as polyacetylene, polythiophene or polypyrrole conduct with $T > 1$ when a few electrons are removed or added to introduce electronic states in the HOMO–LUMO gap of the molecule generally in resonance with the electrode Fermi level, at low bias voltage. Using a perpendicular electric field, those electrons can be driven from one molecular chain to another (molecule **14** in Fig. 1e), acting as a transistor source–drain channel⁶⁶. Synthetic work has produced a number of spiro-bridge compounds (such as **14**)⁶⁷ that confirm the presence of an electronic double-well potential at the bridge site⁶⁸. As in many designs employing electromechanical, electrical-field or even electronic interference effects⁶⁹, the third electrode, which gates the HME transistor, is mesoscopic and not readily scalable to nanoscale dimensions. Coulomb-blockade devices¹⁶ and molecular junctions consisting of two different molecular components^{35,47} have similar scalability problems.

Architecture of hybrid molecular electronic circuits

Given the proposed mode of operation of HME devices and reliability issues, circuit design is emerging as a crucial aspect in the development of future molecular electronic systems, with input/output interconnects, clock frequency and the increase of logic complexity constituting particularly challenging problems. Although NEMS technology may represent a potential solution to reduce the area occupied by the interconnection pads on a wafer, the bandwidth of these interconnects is limited to a few megahertz. The potential switching rate of the molecular component in an HME device, in contrast, is of the order of several terahertz^{15,63}, but it may require optical interconnects, such as subwavelength photonic waveguides⁶¹, to use these frequencies.

Continuation of the ‘top-down’ approach indicates that a conservative architecture with metallic wires interconnecting many devices to form a complex network is feasible using well-defined and spatially localized logic gates, registers and memory cells (Fig. 4). In this case, high-gain three-terminal HME devices are essential for

fan-out (the ability of a given device to provide sufficient power to operate multiple devices interconnected to it) to reduce the wiring complexity between the gates and to lower the power dissipation per gate. The extreme fabrication constraints, especially for complex circuits, will require new architecture paradigms. For instance, it has been proposed to replace process precision by software precision⁷⁰, by using an external nonmolecular processor to select the working HME devices from a massive array of HME devices connected in a checkerboard pattern and to configure a particular circuit. In such cases gain might be localized elsewhere, using traditional CMOS (complementary metal oxide semiconductors) outside the molecular electronic circuitry, with some trade-off in space. Other architectures that have been proposed are quantum cellular automata^{55,71} and quantum computing using a molecule as a quantum dot⁷².

Mono-molecular integration

An ultimate limit of downscaling a circuit made of metallic wires interconnecting HME devices is a circuit where the wires and devices are no longer distinguishable as localized discrete elements of the circuit. This occurs when the interdevice wire length is well below the electronic phase coherence length of the wire material, so that the circuit operates in a ballistic regime where a minute difference between device positioning or wire length will render the circuit unreliable⁷³. Another limit, also experienced in integrated photonics⁷⁴ and mesoscopic low-temperature devices⁷⁵, is that true intrinsic three-terminal devices with external grids are difficult to design in a ballistic regime. A similar problem arises when a metallic control electrode is used to gate a single HME device (Fig. 2). Therefore, the scaling ability of the HME approach is limited even if SWCNTs are used to replace the metallic nanowires for interdevice wiring. Semiconductor-based electronics, when faced at the end of the 1950s with similar limitations that were based on different physical constraints, turned to monolithic technology, which integrates the wiring, transistors and the required passive elements on a single piece of silicon. The conceptually comparable mono-molecular approach (MME) was suggested for molecular electronics as early as the 1980s³⁵, but has only recently received serious

theoretical⁴⁰ and experimental consideration^{45,76}. Tunnelling is one way to overcome the limitations associated with the ballistic transport regime. In this case, electrons are channelled from one reservoir to the other via a network of tunnelling pathways defined by atomic or molecular wires.

The tunnel transparency T between two metallic electrodes is known to exhibit a sharp exponential decay, $T = T_0 e^{-\gamma L}$ with length L through a vacuum gap and $\gamma \approx 2 \text{ \AA}^{-1}$ (ref. 77). Low-gap semiconductor⁷⁸ or molecular materials⁹ tend to reduce the inverse decay length γ . It has been shown theoretically⁷⁹ that tunnelling currents as high as 10 pA under a 0.1-V bias voltage can be driven through specifically designed molecular wires having lengths exceeding 10 nm and a HOMO–LUMO gap in the 1-eV range. Recently, a value of $\gamma = 0.4 \text{ \AA}^{-1}$ was reported for the molecular wire **16** (Fig. 1a), equipped with supporting molecular groups (legs) that exhibit very low T that protect the conjugated wire from a Cu(100) surface. One end of the molecular wire was interconnected to a double atomic step edge and a STM tip used to explore the length dependence of its transparency⁷⁶, confirming that tunnelling electrons can be guided by long molecular wires with an effective section well below 1 nm^2 . On the electronic transport regime map (Fig. 3), this ‘super tunnel’-like phenomenon corresponds to a wire with a section smaller than the de Broglie wavelength of the electron. It can be interpreted as a permanently driven super-exchange electron-transfer mechanism. Although large-scale theoretical and chemical synthesis efforts are required to model and explore this regime, theoretical considerations indicate that values of γ down to $0.05 \text{ \AA}^{-1}$ are achievable for a HOMO–LUMO gap in the 1-eV range⁷⁹. Currently, the chemical synthesis strategies for molecular wires continue to place an emphasis on the length (Fig. 4), such as in **15** (Fig. 1a), rather than on the optimization of their γ values.

Many calculations have explored the possibility of optimizing the electronic contacts between the end of a molecular wire and a metallic electrode, as has been realized in the case of molecule **16** (ref. 76), by investigating the influence of the wire’s end group^{33,36,37}. The metal–molecule contact results from a mixing of the end part of the wire’s molecular orbitals (HOMO, LUMO and the others

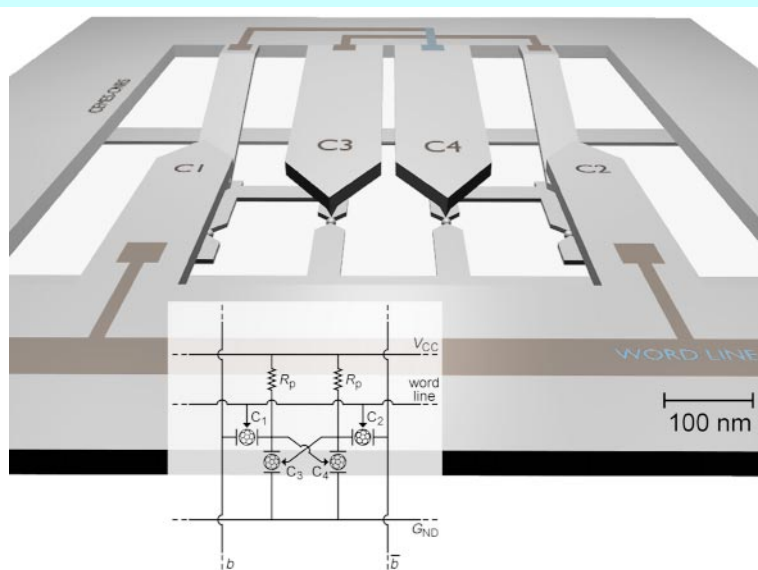


Figure 4 Representative example design of a hybrid molecular electronic device. The figure shows the layout of a memory cell that consists of four individual C_{60} electromechanical transistors⁶⁴ (two for the trigger and two for the driving transistors) under a bias voltage V_{CC} with respect to ground, G_{ND} . The top part is made of four nanoelectromechanics (NEMS) grids plus two metallic NEMS cantilevers, and the lower part is the electrical polarization circuit with polarization resistances R_p made using a metallic constriction along the mesoscopic metallic wires of the circuit (see inset). The full circuit was simulated and optimized using an equivalent electrical circuit for each cantilever and for each C_{60} molecule using the known C_{60} experimental electrical characteristics upon compression¹⁵. Inset, the equivalent electrical circuit diagram. b and $\bar{b}$ carry the digital information to be stored in a memory cell, and the word line activates a line of memory cells for a storage operation.

supporting the tunnel path) with the surface metallic orbitals of the contacts. The resulting extension of the metallic wavefunction of the electrode into the molecular wire may be viewed as a doping effect³⁷ without introducing states in the HOMO–LUMO gap. In the STM images of molecules adsorbed at steps, the apparent height of the molecule is a clear measure of the quality of the electronic contact with their metallic step, which acts as an electrode. This accounts for the pre-factor T_0 in the expression for the tunnel transparency and provides a quantitative criterion for optimizing the metal–molecule contact.

Intramolecular circuits can be formed without metallic pads by using molecular wires. Contrary to conventional electrical circuits, where adding one branch does not change the electronic properties of the others, any new molecular wire or branch added to a molecule effectively creates a ‘new’ molecule with a different electronic structure. Therefore, a standard circuit analysis resulting from the application of Ohm’s and Kirchhoff’s laws is inapplicable even if the design of molecular logic continues to be based on those laws⁸⁰. From elastic scattering theory, the chemical binding of two oligomers with low transparencies T_1 and T_2 in series gives a total transparency T proportional to $T_1 \times T_2$. The binding of two or more oligomers in parallel requires two auxiliary branches to form the nodes (Fig. 6). In this case and when $T < 1$ for each branch, the superposition law yields $T = T_h + T_b + 2(T_h \times T_b)^{1/2}$ (ref. 40) with T_h and T_b being the transparency of each complete individual branch (Fig. 5). However, it is not yet clear whether the invalidity of standard circuit analysis derives solely from the inapplicability of Ohm’s law, emphasizing the need for experimental investigations of intramolecular currents. Nanopore measurements on molecule 17 (Fig. 1e), which consists of three conjugated branches associated in series via two non-conjugated CH_2 nodes⁸⁰, have revealed a very small NDR effect attributed to the central phenyl ring, indicating a multiplicative rather than additive superposition for an association in series. In order to probe along each branch of an intramolecular circuit, large and multi-branched macromolecules⁸¹ need to be synthesized (and imaged), indicating that the fabrication precision

in MME can then be combined by chemical synthesis and nanofabrication techniques, a process referred to as directed self-assembly.

Intramolecular devices will need to be controlled internally without crosstalk to adjacent molecules. Suitable intramolecular wiring and circuitry^{55,80,82} has been proposed, but an intrinsic gating effect that allows internal control has not yet been identified. Even SWCNTs crossing each other⁸³ or junctions⁴⁵ on the molecular scale are typically gated with an electrode connected to a macroscopic pad, suggesting that intramolecular electronic wavefunctions might need to be manipulated differently, such as is done in quantum computing to perform computation inside a single molecule. However, the laws governing the physics of an intramolecular circuit in the tunnelling regime are poorly known. Although it might be possible, for example, to use the vibronic pseudo-continuum associated with many electronic states of a large molecule to break microreversibility and thus design intrinsic three-terminal devices, we cannot yet test this idea directly owing to the difficulties in observing inelastic effects experimentally⁸⁴. Electron correlation effects have now been observed in MWCNTs³⁰ and SWCNTs⁸⁵, but further advances in this direction are required to determine whether a full mono-molecular approach is technologically viable and thus able to provide a realistic alternative to monolithic silicon devices.

Nanofabrication

The fabrication of a circuit using HME devices or the interconnects of MME devices requires the following steps: (1) fabrication of millions of multi-electrode nanojunctions with interelectrode distances down to a few nanometres; (2) wiring these junctions to interconnects; (3) deposition/assembly of one molecule or super-molecule per nanojunction; (4) fabrication of the input/output and driving power interconnects; and (5) packaging the circuit. Currently, two-electrode nanojunctions down to 5 nm have been produced by electron-beam nanolithography, and electrodeposition techniques have revealed that 2 nm can be reached (Fig. 5). Two nearby SWCNTs have also been used to define a nanojunction using

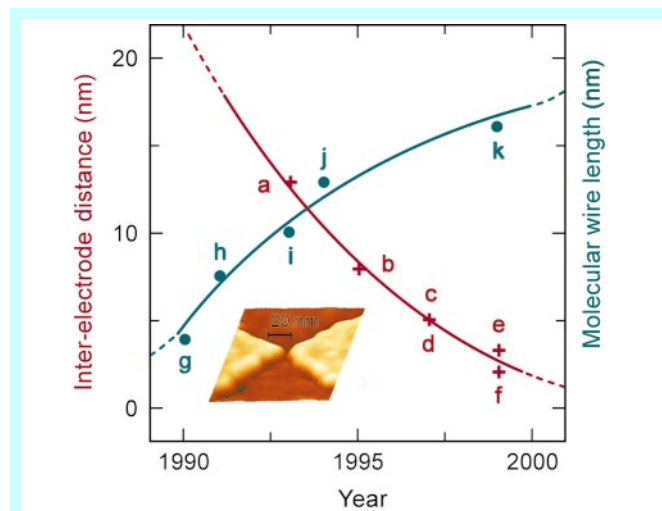


Figure 5 Variation of the inter-electrode distance of planar nanojunctions and the length of the synthesized molecular wires as a function of time. Red, inter-electrode distance; green, molecular wire length. **a, b**, Planar standard e-beam nanolithography nanojunctions, ref. 97 (**a**) and ref. 22 (**b**). **c, e**, Co-planar e-beam nanojunctions with the electrode metal at the same height as the substrate, ref. 21 (**b**) and ref. 27 (**e**). **d, f**, Suspended planar nanojunctions fabricated by electrodeposition⁹⁸ (**d**) and electrochemistry⁹⁹ (**f**). **g, h**, Long oligoimide molecular lines, ref. 100 (**g**) and ref. 101 (**h**). **i**, Thiophene-ethynylene oligomers¹⁰². **j**, Phenylalkyne oligomers¹⁰³. **k**, Alternating block co-oligomers of (1,4-phenylene ethynylene)s and (2,3-thiophene ethynylene)s¹⁰⁴. Inset, AFM tapping mode image of a 4-nm coplanar nanojunction. In **e**, the two gold-palladium metal electrodes are only 1 nm higher than the background silicon wafer surface²⁷.

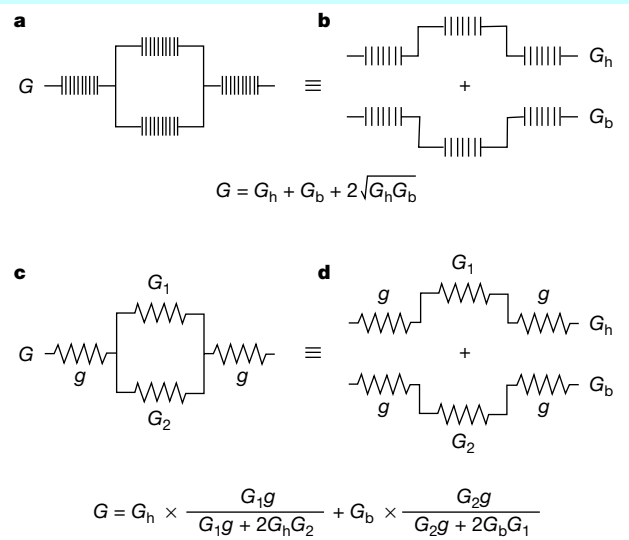


Figure 6 Illustration of the superposition rule operating in a pure tunnelling regime for a simple intramolecular circuit. **a**, The circuit is made of a molecule composed of two branches bonded in parallel and connected to the pad through molecular wires. **b**, The full conductance G of the metal–molecule–metal junction is obtained by decomposing the circuit into two parts having junction conductances G_h and G_b , respectively, and summing up G_h and G_b as indicated (the same rule applies for the transparencies). **c**, By comparison, the conductance G of the same circuit, but with conventional ohmic resistances g , G_1 and G_2 , can also be decomposed in this way (**d**). Of course, this decomposition is never used in circuits, and G for **c** is better calculated using standard circuit analysis.

templates²⁹. For interconnect wiring, a number of approaches exist that go beyond current ultraviolet (UV) or electron-beam lithography. For example, wiring can be nano-imprinted with a resolution limit of more than 10 nm⁸⁶, realized with SWCNTs, or “nanostencilled”⁸⁷ in high vacuum by using movable shadow masks and apertures in cantilevers or a thin silicon membrane when depositing the metallic wires from a source of metal atoms and/or molecules. In addition to being a resist-free, ultraclean nanolithography technique, nanostencilling is a parallel process that allows the simultaneous use of several thousand apertures. Arrays of electron-beam microcolumns are also being developed for parallel nanolithography technology with a resolution of about 10 nm (ref. 88). The array provide low-voltage electron-beams and increase the throughput, with one microcolumn being responsible for each HME circuit or MME circuit interconnection. The accurate placement of molecules in the appropriate position and orientation to form a device, or building of a small circuit with a few SWNTs⁸⁹, is readily achieved through molecular manipulation using scanning-probe microscopy (SPM). But turning this approach into a parallel process will most probably involve patterning of the molecules on surfaces using ‘wet’ methods, such as self-assembly, combing, cantilever techniques, nanostamping or nanostencilling. Self-assembly is a particularly attractive strategy because it could, in principle, exploit thermodynamic control to achieve high precision⁹⁰, but the overall precision with which complex structures can be assembled through molecular self-organization remains to be investigated. Currently, molecular combing can reach a 40% rate of success in connecting SWCNTs⁹¹ between two electrodes. It relies on the force exerted by a liquid meniscus on the SWCNT adsorbed on the wafer and requires a chemical preparation of the wafer surface, a process accessible to nano-imprinting, but not yet applied to oligomers. Nanostamping uses elastomeric stamps made with a pattern of reliefs on their surface. After ‘inking’—a term used to describe the coating of the stamp with molecules—the stamp is contacted to a surface, which allows the accurate reproduction of their area of contact by leaving behind a patterned molecular monolayer in a manner similar to printing^{105,106}. It can attain a precision of a few nanometres but requires that surface and gas phase diffusion be controlled. Shadow masks are well adapted for ultraclean processes⁸⁷, but suitable molecular sources need to be further developed to take into account the MME circuit requirement towards higher molecular weights. Experiments have already explored the use of nonthermal sublimation methods such as matrix-assisted laser desorption (MALD)⁹² or injection of aqueous solutions of molecules into ultrahigh vacuum⁸¹. The ultimate limits of fabrication will probably require the development of “directed self-assembly” methods, where a finite number of molecules or atoms can be assembled and deposited at predefined locations⁹³. Self-assembly, directed by molecular recognition, provides a very efficient means to construct helices, and rack, ladder and grid structures⁹⁴, and provided suitable assembly and electronic functions are “pre-installed” into a molecule during synthesis, it may be used for the self-positioning of HME devices or MME circuits.

Quantum machines

The first proposals for molecular electronics appeared in the 1970s, but it is only the appearance of a number of scientific and economic developments that has allowed the recent resurgence of activity in this field. Crucial are advances in nanoscale science and technology, such as new fabrication methods and probes, which enable individual molecules or small numbers of molecules to be connected in a controlled manner into actual test devices. The driving force behind this research is clearly the need for suitable alternative technologies to Si-based CMOS, which is expected to reach its limitations in 10–20 years⁷¹. Our understanding and concepts of what molecular electronics might look like in the future have changed significantly as a result of all of these factors. In particular, electrode effects and

the extreme sensitivity of quantum tunnelling to small atomic motions, and switching behaviour and rotation⁹⁵ in single molecules, are good examples of possible new operational methodologies that have appeared in the past five years. Finally, the understanding and control of directed self-assembly techniques, combined with developments in synthetic and supramolecular chemistry, increasingly focus on the notion that molecular devices and machines can be synthesized from first principles. The logical progression of this approach is mono-molecular electronics, which integrates many electronic elements, such as wires, switches, and amplifiers, in a single molecule or individual supramolecular assembly. There are many unresolved issues and opportunities for molecular electronic devices in the areas of architecture where a conventional analysis of a molecular device in terms of the linear superposition of its components appears to be invalid. This may make molecular electronics a prime contributor to the building of nanoscale quantum machines using the power of quantum-state superposition. Indeed, as synergistic and cooperative effects in complex molecular systems are understood and controlled, new electronic devices based on unconventional operating principles might be developed. In short, it seems that applying conventional electronic concepts and approaches will not necessarily be the best way to achieve functional molecular devices operating on a dimensional scale where quantum mechanics dominates. □

1. Aviram, A. & Ratner, M. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
2. Riordan, M. & Hodgeson, L. *Crystal Fire: The Birth of the Computer Age* (W. W. Norton & Company, New York, 1997).
3. Taube, H. The electron transfer between metal complexes: a retrospective. *Science* **226**, 1028–1036 (1984).
4. Patoux, C. *et al.* Long-range electronic coupling in bis(cyclometalated) ruthenium complexes. *J. Am. Chem. Soc.* **120**, 3717–3725 (1998).
5. Davis, W. B., Svec, W. A., Ratner, M. A. & Wasielewski, M. R. Molecular-wire behaviour in p-phenylenevinylene oligomers. *Nature* **396**, 60–63 (1998).
6. Frayse, S., Coudret, C. & Launay, J.-P. Synthesis and properties of dinuclear complexes with a photochromic bridge: switching “ON” and “OFF” an intervalence electron transfer. *Eur. J. Inorg. Chem.* 1581–1590 (2000).
7. Patoux, C., Coudret, C., Launay, J.-P., Joachim, C. & Gourdon, A. Topological effects on intramolecular electron transfer via quantum interference. *Inorg. Chem.* **36**, 5037–5049 (1997).
8. Delamarche, E., Michel, B., Biebuyck, H. A. & Gerber, C. Golden interfaces: the surface of self-assembled monolayers. *Adv. Mater.* **8**, 719–729 (1996).
9. Mann, B. & Kuhn, H. Tunnelling through fatty acid salt monolayers. *J. Appl. Phys.* **42**, 4398–4405 (1971).
10. Geddes, N. J., Sambles, J. R., Davis, D. J., Parker, W. G. & Sandman, D. J. Fabrication and investigation of asymmetric current-voltage characteristics of a metal/Langmuir-Blodgett monolayer/metal structure. *Appl. Phys. Lett.* **56**, 1916–1918 (1990).
11. Metzger, R. M. *et al.* Unimolecular electrical rectification in hexadecylquinolinium tricyanoquinodimethanide. *J. Am. Chem. Soc.* **119**, 10455–10466 (1997).
12. Gimzewski, J. K., Stoll, E. P. & Schlittler, R. R. Scanning tunnelling microscopy on individual molecules of copper phthalocyanine adsorbed on polycrystalline silver surfaces. *Surf. Sci.* **181**, 267–277 (1987).
13. Eigler, D. M., Lutz, C. P. & Rudge, W. E. An atomic switch realised with the scanning tunnelling microscope. *Nature* **352**, 600–603 (1991).
14. Joachim, C., Gimzewski, J. K., Schlittler, R. R. & Chavy, C. Electronic transparency of a single C₆₀ molecule. *Phys. Rev. Lett.* **74**, 2102–2105 (1995).
15. Joachim, C. & Gimzewski, J. K. An electromechanical amplifier using a single molecule. *Chem. Phys. Lett.* **265**, 353–357 (1997).
16. Dorogi, M., Gomez, J., Osifchin, R., Andres, R. P. & Reifenger, R. Room-temperature Coulomb blockade from a self-assembled molecular nanostructure. *Phys. Rev. B* **52**, 9071–9077 (1995).
17. Reed, M. A. *et al.* The electrical measurement of molecular junctions. *Ann. NY Acad. Sci.* **852**, 133–144 (1998).
18. Reed, M. A., Zhou, C., Muller, C. J., Burgin, T. P. & Tour, J. M. Conductance of a molecular junction. *Science* **278**, 252–254 (1997).
19. Kerguelis, C. *et al.* Electron transport through a metal-molecule-metal junction. *Phys. Rev. B* **59**, 12505–12513 (1999).
20. Bezryadin, A., Dekker, C. & Schmid, G. Electrostatic trapping of single conducting nanoparticles between nanoelectrodes. *Appl. Phys. Lett.* **71**, 1273–1275 (1997).
21. Rousset, V., Joachim, C., Rousset, B. & Fabre, N. Fabrication of co-planar metal-insulator-metal nanojunction with a gap lower than 10 nm. *J. Phys. III* **5**, 1983–1989 (1995).
22. Di Fabrizio, E. *et al.* Fabrication of 5 nm resolution electrodes for molecular devices by means of electron beam lithography. *Jpn. J. Appl. Phys.* **36**, L70–L72 (1997).
23. Yanson, A. I., Yanson, I. K. & van Ruitenbeek, J. M. Observation of shell structure in sodium nanowires. *Nature* **400**, 144–146 (1999).
24. Porath, D., Bezryadin, A., de Vries, S. & Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **403**, 635–638 (2000).
25. Tans, S. J. *et al.* Individual single wall carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
26. Ebbsen, T. W. *et al.* Electrical conductivity of individual carbon nanotubes. *Nature* **382**, 54–56 (1996).

27. Cholet, S., Joachim, C., Martinez, J. P. & Rousset, B. Fabrication of co-planar metal-insulator-metal nanojunction down to 5 nm. *Eur. Phys. J. Appl. Phys.* **8**, 139–145 (1999).
28. Bachtold, A. *et al.* Contacting carbon nanotubes selectively with low Ohmic contact for four-probe electric measurement. *Appl. Phys. Lett.* **73**, 274–276 (1998).
29. Kong, J., Soh, H. T., Cassell, A. M., Quate, C. F. & Dai, H. Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers. *Nature* **395**, 878–881 (1998).
30. Avouris, Ph. *et al.* in *Proceedings of "Nanotube 1999"* (Plenum, New York, in the press).
31. Mujica, V., Kemp, M., Roitberg, A. & Ratner, M. A. Current-voltage characteristics of molecular wires: eigenvalue staircase, Coulomb blockade and rectification. *J. Chem. Phys.* **104**, 7296–7305 (1996).
32. Porath, D. & Milo, O. Single electron tunnelling and level spectroscopy of isolated C_{60} molecules. *J. Appl. Phys.* **81**, 2241–2244 (1997).
33. Magoga, M. & Joachim, C. Conductance and transparency of long molecular wires. *Phys. Rev. B* **56**, 4722–4729 (1997).
34. Samanta, M. P., Tian, W., Datta, S., Henderson, J. I. & Kubiak, C. P. Electronic conduction through organic molecules. *Phys. Rev. B* **53**, R7626–R7629 (1996).
35. Olson, M. *et al.* A conformation study of the influence of vibration conduction in molecular wires. *Phys. Chem. B* **102**, 941–947 (1998).
36. Yaliraki, S. N., Kemp, M. & Ratner, M. A. Conductance of molecular wires: influence of molecule-electrode binding. *J. Am. Chem. Soc.* **121**, 3428–3434 (1999).
37. Di Ventra, M., Pantelides, S. T. & Lang, N. D. First-principles calculation of transport properties of a molecular device. *Phys. Rev. Lett.* **84**, 979–982 (2000).
38. Joachim, C. & Vinuesa, J. Length dependence of the transparency (conductance) of a molecular wire. *Europhys. Lett.* **33**, 635–640 (1996).
39. Lang, N. D. & Avouris, Ph. Oscillatory conductance of carbon-atom wires. *Phys. Rev. Lett.* **81**, 3515–3518 (1998).
40. Magoga, M. & Joachim, C. Conductance of molecular wires connected or bonded in parallel. *Phys. Rev. B* **59**, 16011–16020 (1999).
41. Lang, N. D. Resistance of atomic wires. *Phys. Rev. B* **52**, 5335–5342 (1995).
42. Ness, H. & Fisher, A. J. Non-perturbative evaluation of STM tunneling probabilities from ab-initio calculations. *Phys. Rev. B* **56**, 12462–12481 (1997).
43. Collins, P. G., Zetti, A., Bando, H., Thess, A. & Smalley, R. E. Nanotube nanodevices. *Science* **278**, 100–103 (1997).
44. Hu, J. T., Min, O. Y., Yang, P. D. & Lieber, C. M. Controlled growth and electrical properties of heterojunctions of carbon nanotubes and silicon nanowires. *Nature* **399**, 48–51 (1999).
45. Yao, Z., Postman, H. W. Ch., Balents, L. & Dekker, C. Carbon nanotube intramolecular junctions. *Nature* **402**, 273–276 (1999).
46. Léonard, F. & Tersoff, J. Novel length scales in nanotube devices. *Phys. Rev. Lett.* **83**, 5174–5177 (1999).
47. Fischer, C. M., Burghard, M., Roth, S. & von Klitzing, K. Organic quantum wells: molecular rectification and single electron tunnelling. *Europhys. Lett.* **28**, 129–134 (1994).
48. Sessoli, R., Gatteschi, D., Caneschi, A. & Novak, M. A. Magnetic bistability in a metal-ion cluster. *Nature* **365**, 141–143 (1993).
49. Manoharan, H. G., Lutz, C. P. & Eigler, D. M. Quantum mirages formed by coherent projection of electronic structure. *Nature* **403**, 512–515 (2000).
50. Sangregorio, S., Ohm, T., Paulsen, C., Sessoli, R. & Gatteschi, D. Quantum tunnelling of the magnetization in an iron cluster nanomagnet. *Phys. Rev. Lett.* **78**, 4645–4648 (1997).
51. Kahn, O. & Launay, J. P. Molecular bistability: An overview. *Chemtronics* **3**, 140–151 (1988).
52. Mathews, R. H. *et al.* A new RTD-FET logic family. *Proc. IEEE* **87**, 596–605 (1999).
53. Gao, H. J. *et al.* Reversible, nanometer-scale conductance transitions in an organic complex. *Phys. Rev. Lett.* **84**, 1780–1783 (2000).
54. Chen, J., Reed, M. A., Rawlett, A. M. & Tour, J. M. Large on-off ratios and negative differential resistance in a molecular electronic device. *Science* **286**, 1550–1552 (1999).
55. Carter, E. L. The molecular device computer: point of departure for large scale cellular automata. *Physica D* **10**, 175–194 (1984).
56. Higelin, D. & Sixl, H. Spectroscopystudies of the photochromism of N-salicylideneaniline mixed crystals and glasses. *Chem. Phys.* **77**, 391–396 (1983).
57. Joachim, C. & Launay, J. P. Bloch effective Hamiltonian for the possibility of molecular switching in the ruthenium-bipyridylbutadiene-ruthenium system. *Chem. Phys.* **109**, 93–99 (1986).
58. Hush, N. S., Wong, A. T., Backs, G. B. & Riemers, J. R. Electron and energy transfer through bridged systems: VI. molecular switches. *J. Am. Chem. Soc.* **112**, 4192–4197 (1990).
59. Gilat, S. L., Kawai, H. S. & Lehn, J. M. Light-triggered electrical and optical switching devices. *J. Chem. Soc. Chem. Commun.* 1439–1442 (1993).
60. Bissel, R. A., Cordova, E., Kaifer, A. E. & Stoddart, J. F. A chemically and electrochemically switchable molecular shuttle. *Nature* **369**, 133–137 (1994).
61. Girard, C., Dereux, A. & Joachim, C. Photonic transfer through subwavelength optical waveguide. *Europhys. Lett.* **44**, 686–692 (1998).
62. Goldhaber-Gordon, D., Montemero, M. S., Love, J. C., Opiteck, G. J. & Ellenbogen, J. C. Overview of nanoelectronic devices. *Proc. IEEE* **85**, 521–539 (1997).
63. Tans, S. J., Verschuere, A. R. M. & Dekker, C. Room temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
64. Joachim, C., Gimzewski, J. K. & Tang, H. Physical principles of the single C_{60} transistor effect. *Phys. Rev. B* **58**, 16407–16417 (1998).
65. Brugger, J., Beljakovic, G., Despont, M., de Rooij, N. F. & Vettiger, P. Silicon micro/nanomechanical device fabrication based on focused ion beam surface modification and KOH etching. *J. Microelec. Eng.* **35**, 401–404 (1997).
66. Aviram, A. Molecules for memory, logic and amplification. *J. Am. Chem. Soc.* **110**, 5687–5692 (1988).
67. Tour, J. M., Rulian, W. & Schumm, J. S. Extended orthogonally fused conducting oligomers for molecular electronic devices. *J. Am. Chem. Soc.* **113**, 7064–7066 (1991).
68. Diers, J. R. *et al.* ESR characterisation of oligomeric thiophene materials. *Chem. Mater.* **6**, 327–332 (1994).
69. Treboux, G., Lapstun, P. & Silverbrook, K. Conductance in nanotube Y-junctions. *Chem. Phys. Lett.* **306**, 402–406 (1999).
70. Collier, C. P. *et al.* Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
71. Compagno, R., Molenkamp, L. & Paul, D. J. (eds) *Technology Roadmap for European Nanoelectronics* (European Commission, IST Program, Brussels, 1999).
72. Nakamura, Y., Pashkin, Yu. A. & Tsai, T. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
73. Landauer, R. Can we switch by control of quantum mechanical transmission? *Phys. Today* 119–121 (1989).
74. Keyes, R. W. Lighting up logic. *Nature* **362**, 289–290 (1993).
75. Washburn, S., Schmid, H., Kern, D. & Webb, R. A. Normal-metal Aharonov-Bohm effect in the presence of a transverse electric field. *Phys. Rev. Lett.* **59**, 1791–1794 (1987).
76. Langlais, V. *et al.* Spatially resolved tunnelling along a molecular wire. *Phys. Rev. Lett.* **83**, 2809–2812 (1999).
77. Simmons, J. G. Generalized formula for electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* **34**, 1793–1803 (1963).
78. Lewicki, G. & Mead, C. A. Experimental determination of the E-k relationship in electron tunnelling. *Phys. Rev. Lett.* **16**, 939–941 (1966).
79. Magoga, M. & Joachim, C. Minimal attenuation for tunnelling through a molecular wire. *Phys. Rev. B* **57**, 1820–1823 (1998).
80. Tour, J. M., Kozaki, M. & Seminario, J. M. Molecular scale electronics: a synthetic/computational approach to digital computing. *J. Am. Chem. Soc.* **120**, 8486–8493 (1998).
81. Sugiura, K. I., Tanaka, H., Matsumoto, T., Kawai, T. & Sakata, Y. A mandala-patterned bandanna-shaped porphyrin oligomer, $C_{1244}H_{1350}N_{84}Ni_{20}O_{88}$, having a unique size and geometry. *Chem. Lett.* 1193–1194 (1999).
82. Wada, Y. Atom electronics: a proposal of atom/molecule switching devices. *Ann. NY Acad. Sci.* **852**, 257–276 (1998).
83. Menon, M. & Srivastava, D. Carbon nanotube "T-junctions": nanoscale metal-semiconductor-metal contact devices. *Phys. Rev. Lett.* **79**, 4453–4456 (1997).
84. Stipe, B. C., Rezaei, M. A. & Ho, W. Inducing and viewing the rotation motion of a single molecule. *Science* **279**, 1907–1909 (1998).
85. Tans, S. J., Devoret, M. H., Groeneveld, R. J. A. & Dekker, C. Electron-electron correlations in carbon nanotubes. *Nature* **394**, 761–764 (1998).
86. Guo, L., Krauss, P. R. & Chou, S. Y. Nanoscale silicon field effect transistors fabricated using imprint lithography. *Appl. Phys. Lett.* **71**, 1881–1883 (1997).
87. Lüthi, R. *et al.* Parallel nanodevice fabrication using a combination of shadow mask and scanning probe methods. *Appl. Phys. Lett.* **75**, 1314–1316 (1999).
88. Kratschmer, E. *et al.* An electron-beam microcolumn with improved resolution beam current and stability. *J. Vac. Sci. Technol. B* **13**, 2498–2503 (1995).
89. Lefebvre, J., Lynch, J. F., Llaguno, M., Radosavljevic, M. & Johnson, A. T. Single-wall carbon nanotube circuits assembled with an atomic force microscope. *Appl. Phys. Lett.* **75**, 3014–3016 (1999).
90. Allara, D. L. *et al.* Evolution of strategies for self-assembly and hook up of molecule-based devices. *Ann. NY Acad. Sci.* **852**, 349–370 (1998).
91. Gerdes, S., Ondarcuhu, T., Cholet, S. & Joachim, C. Combining a carbon nanotube on a flat metal-insulator-metal nanojunction. *Europhys. Lett.* **48**, 292–298 (1999).
92. M. Handschuh, M., Nettesheim, S. & Zenobi, R. *Appl. Surf. Sci.* **137**, 125–135 (1999).
93. Philp, D. & Stoddart, J. F. Self-assembly in natural and unnatural systems. *Angew. Chem. Int. Edn Engl.* **35**, 1154–1196 (1996).
94. Drain, C. M. & Lehn, J. M. Self-assembly of square multiporphyrin arrays by metal ion coordination. *J. Chem. Soc. Chem. Commun.* 2313–2315 (1994).
95. Gimzewski, J. K. & Joachim, C. Nanoscale science of single molecules using local probes. *Science* **283**, 1683–1688 (1999).
96. Yazdani, A., Eigler, D. M. & Lang, N. D. Off-resonance conduction through atomic wires. *Science* **272**, 1912–1924 (1996).
97. Fischer, P. B. & Chou, S. Y. 10 nm electron beam lithography and sub-50 nm overlay using a modified scanning electron microscope. *Appl. Phys. Lett.* **62**, 2989–2991 (1993).
98. Bezryadin, A. & Dekker, C. Nanofabrication of electrodes with sub-5 nm spacing for transport experiments on single molecules and metal clusters. *J. Vac. Sci. Technol. A* **15**, 793–799 (1997).
99. Morpurgo, A. E., Marcus, C. M. & Robinson, D. B. Controlled fabrication of metallic electrodes with atomic separations. *Appl. Phys. Lett.* **74**, 2084–2086 (1999).
100. Dietz, T. M., Stallman, B. J., Kwan, W. S. V., Penneau, J. F. & Miller, L. L. Soluble oligoimide molecular lines which have persistent poly(anion radicals) and poly(dianions). *J. Chem. Soc. Chem. Commun.* 367–369 (1990).
101. Kwan, W. S. V., Atanasoska, L. & Miller, L. L. Oligoimide monolayers covalently attached to gold. *Langmuir* **7**, 1419–1425 (1991).
102. Schumm, J. S., Pearson, D. L. & Tour, J. M. Iterative divergent/convergent doubling approach to linear conjugated oligomers. A rapid route potential molecular wire. *Macromolecules* **27**, 2348–2350 (1994).
103. Schumm, J. S., Pearson, D. L. & Tour, J. M. Iterative divergent/convergent doubling approach to linear conjugated oligomers. A rapid route to a 128 Å long potential molecular wire. *Angew. Chem. Int. Edn Engl.* **33**, 1360–1363 (1994).
104. Huang, S. & Tour, J. M. Rapid solid-phase synthesis of conjugated homo-oligomers and (AB) alternating block co-oligomers of precise length and constitution. *J. Org. Chem.* **64**, 8898–8906 (1999).
105. Kumar, A. & Whitesides, G. M. Features of gold having micrometer to centimeter dimensions can be formed through a combination of stamping with an elastomeric stamp and an alkanethiol 'ink' followed by chemical etching. *Appl. Phys. Lett.* **63**, 2002–2004 (1993).
106. Kumar, A., Biebuyck, H. A. & Whitesides, G. M. Patterning SAMs: Applications in materials science. *Langmuir* **10**, 1498–1511 (1994).

Acknowledgements

We thank the CEMES Molecular Electronics group and IBM Zurich's Science and Technology department for helpful discussions. This work has partially been supported through the European Union and the Swiss Federal Office for Education and Science by the Information Society Technologies—Future Emerging Technology (IST—FET) project Bottom Up Nanomachines.

Correspondence and requests for materials should be addressed to C. J. (e-mail: joachim@cemes.fr).

Superconductivity at 52 K in hole-doped C₆₀

J. H. Schön, Ch. Kloc & B. Batlogg

Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

Superconductivity in electron-doped C₆₀ was first observed almost ten years ago. The metallic state and superconductivity result from the transfer of electrons from alkaline or alkaline-earth ions to the C₆₀ molecule, which is known to be a strong electron acceptor. For this reason, it is very difficult to remove electrons from C₆₀—yet one might expect to see superconductivity at higher temperatures in hole-doped than in electron-doped C₆₀, because of the higher density of electronic states in the valence band than in the conduction band. We have used the technique of gate-induced doping in a field-effect transistor configuration to introduce significant densities of holes into C₆₀. We observe superconductivity over an extended range of hole density, with a smoothly varying transition temperature T_c that peaks at 52 K. By comparison with the well established dependence of T_c on the lattice parameter in electron-doped C₆₀, we anticipate that T_c values significantly in excess of 100 K should be achievable in a suitably expanded, hole-doped C₆₀ lattice.

Superconductivity at relatively high temperatures in metallic C₆₀ has its origin essentially in two favourable materials properties. First, the frequency spectrum of the C₆₀ intramolecular vibrations extends to high energies (~ 200 meV) and several of these vibrations couple significantly to the electronic states, providing intramolecular electron–phonon interaction^{1–5}. Second, the electronic density of states in the crystal is high because the weak overlap of electronic wavefunctions in the (van der Waals bonded) C₆₀ crystals leads to narrow bands. The widths of both the valence and the conduction bands are calculated to be approximately the same (~ 0.5 eV; refs 6, 7); however, the valence-band density of states (DOS) is higher⁶ as it is derived from a fivefold degenerate (h_{1u}) HOMO (highest occupied molecular orbital) state compared to the threefold degenerate (t_{1u}) LUMO (lowest unoccupied molecular orbital) state. Assuming similar or slightly stronger electron–phonon coupling in the valence band (ref. 8; general reasons for enhanced T_c in hole-doped metals have been suggested in ref. 9), we thus expect a higher T_c when holes are introduced into this band. But it is very difficult to oxidize C₆₀ (refs 10, 11), and we are not aware of a successful method of producing metallic bulk C₆₀ doped with holes. (However, we note that when a fullerene mixture was exposed to highly oxidizing interhalogen vapour, magnetic signatures developed (~ 50 p.p.m. of the expected full Meissner signal) that were interpreted as being due to superconductivity in a very small portion of the sample¹².)

A further question concerning the metallic state in C₆₀ can be addressed by our experiments on gate-induced doping using a field-effect transistor (FET) configuration^{13,14} (see Fig. 1 inset): will correlation among electrons lead to a Mott–Hubbard insulating state for particular carrier concentrations? When electrons are doped into C₆₀ by chemical means, the metallic state appears to be realized for specific numbers of dopant atoms per C₆₀ molecule. In particular, the A₃C₆₀ composition is the most favourable for superconductivity, while A₄C₆₀ appears not even to be metallic¹⁵. The electronic structure of doped C₆₀ is therefore under discussion^{16,17}. It is thus of fundamental interest to follow the electronic state while the doping level is continuously varied by means of the applied gate voltage, without the crystallographic changes that are inevitable when dopant atoms are incorporated into the C₆₀ lattice.

We report here our observations of gate-induced superconductivity in hole- and electron-doped C₆₀ single-crystal field-effect transistor structures. We have been able to sweep the charge carrier

density continuously from ~ 4.5 holes per C₆₀ to ~ 4.5 electrons per C₆₀. We observe a maximum T_c of 52 K for 3–3.5 holes per C₆₀, which is almost five times higher than for electron-doping in the same geometry (11 K)¹³ and is to the best of our knowledge the highest T_c reported for a non-copper-oxide superconductor. We also analyse the dependence of the superconductivity on the charge carrier density induced by gate-doping.

Sample preparation

We prepared FETs on C₆₀ crystals grown by physical vapour transport¹⁸. This transport process is an efficient method of purifying C₆₀. A charge of 10 to 20 mg C₆₀ (pre-purified by multiple sublimations) was placed in a quartz tube (both sides open) inside a horizontal furnace and was exposed to a stream of pure hydrogen, flowing at a rate of 40–100 ml min^{−1} at a temperature of 600 °C. Crystals 1–2 mm in size nucleated spontaneously on the reactor

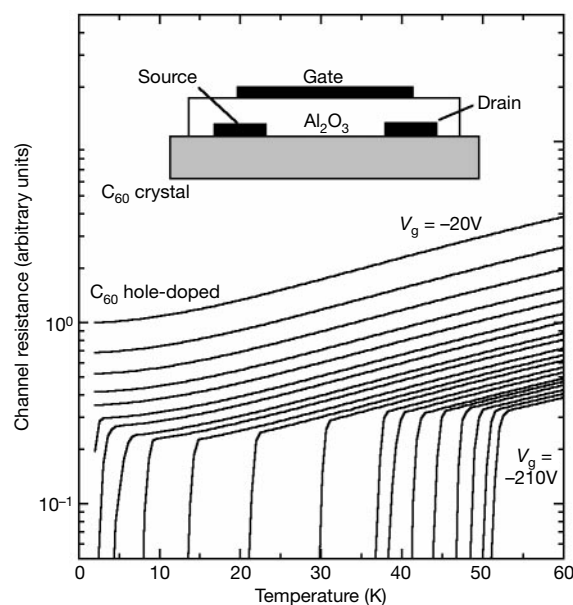


Figure 1 Channel resistance of a C₆₀ single-crystal field-effect transistor. The gate voltage (V_g) was varied between -20 and -210 V in -10 V increments; the shift of the transition temperature T_c , indicated by the break in the slope, with the applied bias is clearly observable. Inset, the field-effect transistor geometry used in the experiments.

wall, and grew stress free to the centre of the reactor. These crystals have faces that are either (100) or (111); we removed these crystals from the reactor and deposited (by evaporation) gold source and drain electrodes, forming a channel of $\sim 25\ \mu\text{m}$ length and $500\text{--}1,000\ \mu\text{m}$ width. An aluminium oxide dielectric layer and a gate electrode complete the FET structure (see Fig. 1 inset). These devices show n- as well as p-channel activity, reflecting the ambipolar transport in these high-quality single crystals and further emphasizing the low interface state density of the FETs¹³. The electrical measurements were performed in a Quantum Design system by sweeping either the temperature (T) or the gate voltage ($T > 1.7\ \text{K}$, magnetic field up to $9\ \text{T}$).

Using the field-effect doping technique, we have demonstrated superconductivity in electron-doped C_{60} single crystals at $11\ \text{K}$ (ref. 13). Moreover, we have used this technique to create a metallic state in organic molecular crystals, such as pentacene, tetracene and anthracene¹⁴. These materials were the first hydrocarbon superconductors¹⁴.

Hole superconductivity

Figure 1 shows the channel resistance as a function of temperature for various negative gate voltages (hole-doping). The resistance systematically decreases as more holes are injected. The decrease is essentially linear in applied voltage, with a small offset at low voltages that might be associated with the presence of localized electronic states. The normal-state resistivities are in the range of a few $\text{m}\Omega\ \text{cm}$, assuming that the injected charge is confined to a single C_{60} molecular layer. The central result of this study is the observation of superconductivity above $50\ \text{K}$ ($T_c \approx 52\ \text{K}$).

In Fig. 2 we plot T_c —which is inferred from the breaks in the slope of the resistivity data—for two different samples versus the number of holes per molecule. The latter is calculated from the independently measured capacitance of the gate dielectric, assuming that all the charge is located in a single layer of molecules. This assumption is meaningful in the light of existing knowledge about FET structures and actual calculations for organic semiconductors¹⁹.

In addition to the high T_c , we also note that superconductivity occurs over a rather wide range of hole concentrations. Furthermore, T_c gradually increases after having reached the experimental lower limit of $1.7\ \text{K}$ near 1 hole per C_{60} , and reaches a maximum between 3 and 3.5 holes per C_{60} . Over the limited range of data from this study, the initial rise of T_c appears to be gradual, suggesting an exponential or power-law dependence (see Fig. 2 inset). Therefore, following T_c to lower concentrations and temperatures, where other

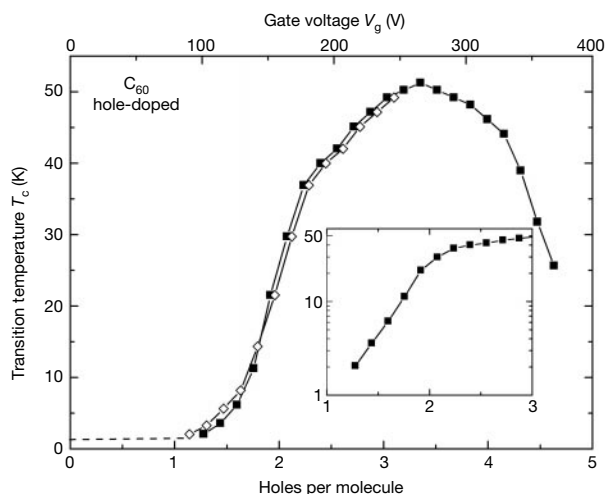


Figure 2 Transition temperature T_c as a function of hole density for two samples. T_c exhibits a maximum of $52\ \text{K}$ at approximately 3.2 holes per C_{60} . The minimum measurable T_c in the experimental set-up ($1.7\ \text{K}$) is indicated by the dashed line. Inset, data replotted on a semi-logarithmic scale.

many-body ground states might occur, is desirable. The continuous increase of T_c with increasing hole concentration appears to be in contrast to the sharp onset of gate-induced superconductivity in hole-doped polyacene single crystals¹². At the highest doping level (~ 4.5 holes per C_{60}), T_c drops to half its maximum value. The two sets of T_c values from different samples are in good agreement, and both suggest a shoulder in the T_c curve near $2\text{--}2.5$ holes per C_{60} , which corresponds to a quarter-filled valence band. Further studies will be needed in order to determine if this feature reflects a structure in the density of states.

The superconducting coherence length ξ , one of the characteristic parameters of superconductivity, can be deduced from the suppression of T_c in a magnetic field H . In Fig. 3 we plot $T_c(H)$ with the field applied perpendicular to the superconducting layer for a carrier concentration that results in a T_c of $41.5\ \text{K}$. The slope of the upper critical field dH_{c2}/dT is $\sim 0.7\ \text{T K}^{-1}$ and the usual extrapolation yields $20\ \text{T}$ for H_{c2} at $T = 0$. This corresponds to a coherence length of $40\ \text{\AA}$, which is similar to the values measured in electron-doped C_{60} (refs 1–3).

T_c versus carrier concentration

In the following discussion of additional data, we try to shed some light on the origin of the high T_c in hole-doped C_{60} . It is helpful to keep in mind that T_c can be varied experimentally by changing the band filling (doping level) or by changing the bandwidth (lattice parameter). We will first discuss our FET data, which are results for a fixed lattice parameter ($\sim 14.16\ \text{\AA}$) and bandwidth. We will then discuss the effect of lattice-parameter variation in A_3C_{60} ; this variation resulted from the insertion of ions (A) of various size by chemical doping.

In Fig. 4a we compare the dependence of T_c on the density of electrons and holes, and we find significant differences. We induced up to almost 5 electrons per C_{60} , and thereby mapped out the dependence of T_c (above $1.7\ \text{K}$) on the electron density. Previous data on chemically doped bulk samples have shown a very similar peaking of T_c near 3 electrons per C_{60} (ref. 20), and thus the superconducting region extends from ~ 2.5 to 3.6 electrons per C_{60} , close to half-filling of the LUMO band. In the case of hole-doping T_c is almost five times higher and, moreover, the superconductivity range is much wider. Furthermore, T_c reaches its maximum of $52\ \text{K}$ at $3\text{--}3.5$ holes per C_{60} , less than the 5 holes per C_{60} which would correspond to half filling of the HOMO bands. At 5 holes per C_{60} , T_c either vanishes or is at least strongly suppressed.

The results in Fig. 4 represent an extremely wide shift of the chemical potential through the electronic states of a solid, corresponding to a continuous variation from a 'half-empty' valence band to a 'half-filled' conduction band. It is thus tempting to compare T_c to the underlying electronic DOS. Figure 4b shows

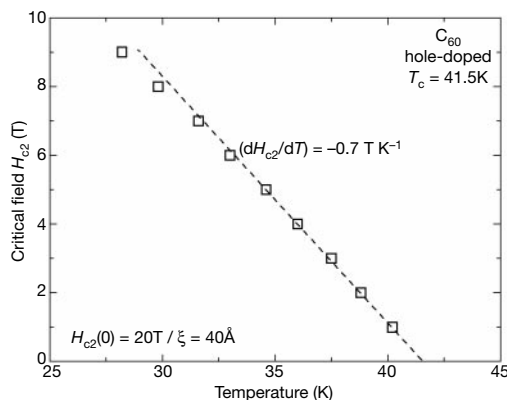


Figure 3 Critical magnetic field H_{c2} as a function of temperature. A superconducting coherence length ξ_0 of $40\ \text{\AA}$ and an upper critical field $H_{c2}(0)$ of $20\ \text{T}$ can be estimated from the temperature dependence of H_{c2} .

the DOS calculated for C_{60} (ref. 6), as a function of induced charge per molecule rather than as a function of energy. This calculation was done for a bulk three-dimensional solid. Our situation is different in that the two-dimensional (2D) metal is sandwiched between the undoped C_{60} bulk and the aluminium oxide gate dielectric. Although it would be highly desirable to repeat the calculations for the present situation, we will simply assume that the essential features of the DOS in Fig. 4b remain relevant.

First, we notice that the DOS on the hole-doping side is indeed greater than on the electron-doping side, as expected for the higher degeneracy of the states. Second, T_c does not bear a close resemblance to the DOS curve. The maxima in T_c occur at doping levels well removed from the DOS maxima. This is particularly noticeable on the hole-doping side, where the DOS peaks strongly near 1 hole per C_{60} , while T_c is below the experimental threshold of 1.7 K. Third, a significant carrier density has to be introduced before superconductivity is observed (above 1.7 K), well after the metallic state is developed, as indicated by the temperature dependence of the resistivity. Further detailed studies of the resistivity near T_c are required to investigate the importance of 2D fluctuations in the gate-induced 2D superconductors. Homogeneous doping by chemical means appears not to be possible for this range¹². Thus, Fig. 4 reveals that the DOS might not be the dominant factor determining T_c , and that T_c is also strongly modified by a filling-dependent coupling strength. This is not unexpected considering that the bands are derived from three and five molecular orbitals, respectively. For the doping range explored here, more than one sub-band is occupied by holes and electrons. The coupling strengths for these sub-bands are obviously not the same.

Coupling strength

Insight into the coupling strength can be gained by examining the resistivity above T_c . Figure 5 shows the resistivity $\rho(T)$ for electron and hole doping, with the hole concentration chosen to produce a T_c similar to the maximum T_c for electron doping, that is, approximately 1.8 holes per C_{60} compared to 3 electrons per C_{60} (see Fig. 4).

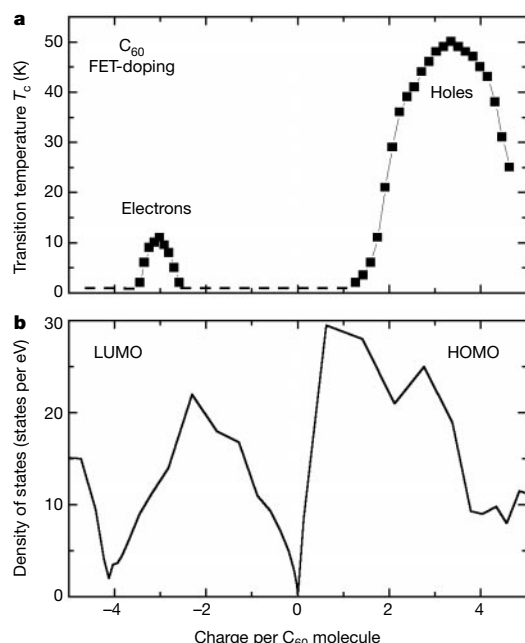


Figure 4 Transition temperature and density of states as a function of charge per C_{60} molecule. **a**, The transition temperature for electron-doped and hole-doped C_{60} in the present FET geometry. In these gate-doped devices, the maximum T_c is 11 K for electrons and 52 K for holes. The broken line indicates the doping range where no superconductivity above 1.7 K was found in this study. **b**, The electronic density of states calculated for three-dimensional bulk C_{60} (ref. 5).

For a quantitative comparison, the measured $\rho(T)$ for the hole-doped metal has been multiplied by a factor 1.8/3 in order to account for the difference in carrier density (both curves are for 3 charge carriers per molecule and can, therefore, also be compared to A_3C_{60}). This linear scaling with carrier concentration is indeed appropriate, as we have found in extensive studies of $\rho(T)$ for various carrier concentrations in this range, which will be discussed elsewhere.

The main result in Fig. 5 is clear: the resistivity of the hole-doped metal is five to six times higher than that of the electron-doped metal, for the same number of charge carriers per molecule. This difference indicates stronger hole–phonon than electron–phonon coupling, and thus the difference in T_c has to be ascribed to differences in coupling strength. We now consider some features of the $\rho(T)$ curves in Fig. 5. Both data sets reflect the phase transition near 250 K (ref. 21) in the form of an approximately 5% reduction in $\rho(T)$. The absolute resistivity values of the FET electron-doped C_{60} can be compared with those of chemically doped C_{60} , either in bulk or in thin-film form^{1,22}. The residual resistivity ρ_0 in the present samples is in the range 250–300 $\mu\Omega$ cm, and is thus about half that of very good chemically doped A_3C_{60} samples. However, the temperature-dependent part of the resistivity $\rho(T) - \rho_0$ is significantly larger in the gate-doped samples, resulting in a variation of $\rho(T) - \rho_0$ from $\leq 20 \mu\Omega$ cm at T_c to $\sim 1.8 \text{ m}\Omega$ cm at room temperature²³ (see Fig. 6). This variation is only $\sim 0.5 \text{ m}\Omega$ cm in chemically doped C_{60} .

The hole-doped metal is different from the electron-doped metal not only in terms of the coupling strength, which is in accordance with theoretical estimations¹⁰, but also in the details of the coupling. This is shown in Fig. 6, where we plot $\rho(T) - \rho_0$ on a logarithmic scale. (Keeping in mind that the measured data are not corrected for possible temperature-induced differential strains and for thermal expansion effects^{24,25}, a comparison is nevertheless meaningful since all experimental parameters are the same for the two curves, except for the sign of the applied gate voltage.) At low temperatures the resistivity of the electron-doped metal varies less rapidly than that of the hole-doped metal. At higher temperatures, however, both types vary in a similar way: $\rho \propto T^n$ with $n \approx 2$. (This is expected from scattering of electron or holes in two dimensions by acoustic phonons in a three-dimensional crystal; C. M. Varma, personal communication.) The two metallic states therefore differ also in the energy dependence of the coupling constant, the hole-doped metal

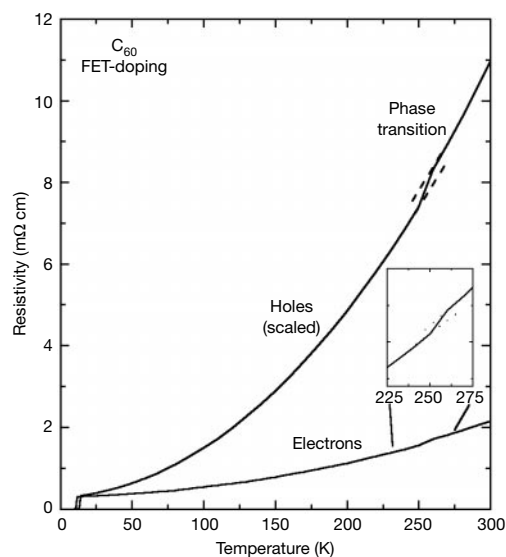


Figure 5 Resistivity of electron- and hole-doped C_{60} as a function of temperature. The resistivity of the hole-doped sample has been scaled (multiplied by ~ 0.6) in order to compare samples of the same carrier density (three charge carriers per molecule). The superconducting phase transition at low temperatures and the structural phase transition around 255 K are clearly observable.

having more weight at low energies. As the resistivity is given by the DOS at the Fermi level, $N(0)$, times the square of the coupling strength, λ^2 , we cannot extract the values of λ independently. If the hole DOS is about twice the electron DOS, as can be estimated for the three-dimensional case, the ratio of the λ values is approximately 1.5, which is reasonable in the light of the T_c ratio. The experimental approach employed here offers an opportunity for future detailed studies of the coupling strength and its temperature (and energy) dependence as the chemical potential is moved through the band structure.

Outlook

The T_c value of 52 K for hole-doped C_{60} is to our knowledge the highest yet reported for a non-copper-oxide superconductor. Furthermore, the lattice parameter a_0 in the FET geometry is essentially that of undoped C_{60} crystals ($a_0 \approx 14.16 \text{ \AA}$). It has been demonstrated^{1–3} that T_c can be substantially raised by expanding the lattice through incorporation of ions. In the present geometry, where the metallic layer is clamped, the gate-induced charge might slightly modify the intramolecular bond lengths. It is

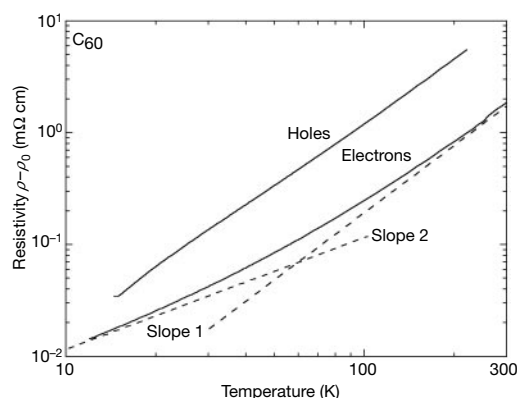


Figure 6 Difference of the resistivity ρ and the residual resistivity ρ_0 as a function of temperature. The dashed lines show a power-law dependence with exponent one and two for comparison.

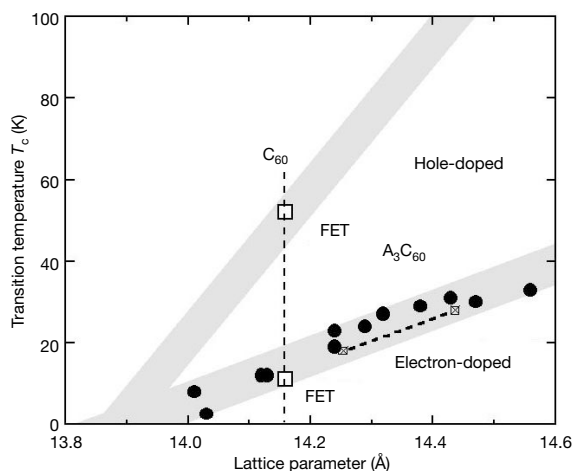


Figure 7 Transition temperature in electron- and hole-doped C_{60} as function of lattice parameter. In A_3C_{60} , with three electrons per C_{60} , the lattice parameter is varied through the size of the inserted A ion (Na, K, Rb, Cs)^{1–3} and extrapolates to a very small value for lattice parameters in the range of 13.8–13.9 Å (filled circles, lower grey line). The dotted line represents the T_c variation when the lattice parameter is reduced by hydrostatic pressure^{1–3}. The lower open square shows T_c when three electrons, as in A_3C_{60} , are introduced in C_{60} crystals through the gate voltage in a field-effect geometry (FET)¹². The upper open square is the present result for FET hole-doped C_{60} . Assuming similar scaling of T_c in hole- and in electron-doped C_{60} , the upper grey line is an attempt to extrapolate T_c for hole-doped C_{60} with an expanded lattice.

unclear to what extent the intermolecular distances and therefore the electronic bandwidth would be affected by this. However, this effect will be significantly smaller than the variation of the intermolecular distances due to the insertion of ions or neutral molecules by chemical doping for a fixed carrier density (see Fig. 7). This observation has been ascribed^{1–3} to the change of bandwidth (that is, DOS) due to variations of the wavefunction overlap between adjacent molecules.

The effect of alkali-metal doping is well studied for A_3C_{60} , and amounts to an increase of T_c from ~ 14 – 15 K at $a_0 \approx 14.16 \text{ \AA}$ to $T_c \approx 33$ K for $a_0 \approx 14.56 \text{ \AA}$ (refs 1–3). To estimate the influence of a similar lattice expansion for the hole-doped situation we attempt a semi-quantitative estimation of T_c (Fig. 7). Underlying this estimation is the experimental result for A_3C_{60} that T_c extrapolates to zero for a lattice parameter of $13.9 \text{ \AA}$, owing to the rapid broadening of the LUMO band leading to a decrease of the DOS. This band broadening will also hold for the HOMO states and, therefore, one might, in a simplified way, draw a straight line between this point and the experimental value of 52 K as an extrapolation of T_c . If lattice expansions similar to those achieved in the past for electron-doping could be realized also with hole-doped C_{60} , we anticipate that values of T_c well above 100 K could be achieved (for $a_0 \approx 14.6 \text{ \AA}$). □

Received 5 September; accepted 30 October 2000.

- Seitz, F. & Turnbull, D. *Solid State Physics* Vol. 48 (eds Ehrenreich, H. & Spaepen, F.) (Academic, San Diego, 1994).
- Ramirez, A. P. C_{60} and its superconductivity. *Supercond. Rev.* **1**, 1–101 (1994).
- Gunnarsson, O. Superconductivity in fullerenes. *Rev. Mod. Phys.* **69**, 575–606 (1997).
- Hebard, A. F. *et al.* Superconductivity at 18 K in potassium-doped C_{60} . *Nature* **350**, 600–601 (1991).
- Varma, C. M., Zaanen, J. & Raghavachari, K. Superconductivity in the fullerenes. *Science* **254**, 989–992 (1991).
- Erwin, S. C. in *Buckminsterfullerenes* (eds Billups, W. E. & Ciufolini, M. A.) 217–255 (VCH, New York, 1992).
- Haddon, R. C., Siegrist, T., Fleming, R. M., Bridenbaugh, P. M. & Laudise, R. A. Band structures of organic thin-film transistor materials. *J. Mater. Chem.* **5**, 1719–1724 (1995).
- Mazin, I. I. *et al.* Quantitative theory of superconductivity in doped C_{60} . *Phys. Rev. B* **45**, 5114–5117 (1992).
- Hirsch, J. E. Bond-charge repulsion and hole superconductivity. *Physica C* **158**, 326–336 (1989).
- Haddon, R. C. The fullerenes: powerful carbon-based electron acceptors. *Phil. Trans. R. Soc. Lond. A* **343**, 53–62 (1993).
- Reed, C. A., Kim, K. C., Bolskar, R. D. & Mueller, L. J. Taming superacids: stabilization of the fullerene cations HC_{60}^+ and C_{60}^+ . *Science* **289**, 101–104 (2000).
- Song, L. W., Fredette, K. T., Chung, D. D. L. & Kao, Y. H. Superconductivity in interhalogen-doped fullerenes. *Solid State Commun.* **87**, 387–391 (1993).
- Schön, J. H., Kloc, Ch., Haddon, R. C. & Batlogg, B. A superconducting field-effect switch. *Science* **288**, 656–658 (2000).
- Schön, J. H., Kloc, Ch. & Batlogg, B. Superconductivity in molecular crystals induced by charge injection. *Nature* **406**, 704–706 (2000).
- Kochanski, G. P., Hebard, A. F., Haddon, R. C. & Fiory, A. T. Electrical resistivity and stoichiometry of K_3C_{60} films. *Science* **255**, 184–186 (1992).
- Rosinsky, M. J. Recent developments in the chemistry and physics of metal fullerenes. *Chem. Mater.* **10**, 2665–2685 (1998).
- Han, J. E., Koch, E. & Gunnarsson, O. Metal-insulator transitions: influence of lattice structure, Jahn-Teller effect, and Hund's Rule coupling. *Phys. Rev. Lett.* **84**, 1276–1279 (2000).
- Kloc, Ch., Simpkins, P. G., Siegrist, T. & Laudise, R. A. Physical vapor growth of centimeter-sized crystals of α -hexathiophene. *J. Cryst. Growth* **182**, 416–427 (1997).
- Dodabalapur, A., Torsi, L. & Katz, H. E. Organic transistors: two-dimensional transport and improved electrical characteristics. *Science* **268**, 270–271 (1995).
- Yildirim, T. *et al.* T_c vs. carrier concentration in cubic fulleride superconductors. *Phys. Rev. Lett.* **77**, 167–170 (1996).
- Heiney, P. A. *et al.* Discontinuous volume change at the orientational-ordering transition in solid C_{60} . *Phys. Rev. B* **45**, 4544–4547 (1992).
- Hesper, R., Tjeng, L. H., Heeres, A. & Sawatzky, G. A. BCS-like density of states in superconducting A_3C_{60} surfaces. *Phys. Rev. Lett.* **85**, 1970–1973 (2000).
- Klein, O., Grüner, G., Huang, S.-M., Wiley, J. B. & Kaner, R. B. Electrical resistivity of K_3C_{60} . *Phys. Rev. B* **46**, 11247–11249 (1992).
- Crespi, V. H., Hou, J. G., Xiang, X.-D., Cohen, M. L. & Zettl, A. Electron-scattering mechanisms in single-crystal K_3C_{60} . *Phys. Rev. B* **46**, 12064–12067 (1992).
- Vareka, W. A. & Zettl, A. Linear temperature dependent resistivity at constant volume in Rb_3C_{60} . *Phys. Rev. Lett.* **72**, 4121–4124 (1994).

Acknowledgements

We thank E. A. Chandross and C. M. Varma for discussions, and E. Bucher for the use of his equipment.

Correspondence and requests for materials should be addressed to B.B. (e-mail: batlogg@lucent.com).

Genetic control and evolution of sexually dimorphic characters in *Drosophila*

Artyom Kopp^{*†}, Ian Duncan[†] & Sean B. Carroll^{*}

^{*} Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706-1596, USA

[†] Department of Biology, Washington University, St Louis, Missouri 63130, USA

Sexually dimorphic abdominal pigmentation and segment morphology evolved recently in the *melanogaster* species group of the fruitfly *Drosophila*. Here we show that these traits are controlled by the *bric-a-brac* (*bab*) gene, which integrates regulatory inputs from the homeotic and sex-determination pathways. *bab* expression is modulated segment- and sex-specifically in sexually dimorphic species, but is uniform in sexually monomorphic species. We suggest that *bab* has an ancestral homeotic function, and that regulatory changes at the *bab* locus played a key role in the evolution of sexual dimorphism. Pigmentation patterns specified by *bab* affect mating preferences, suggesting that sexual selection has contributed to the evolution of *bab* regulation.

A key challenge in evolutionary biology is to identify genetic events responsible for morphological change, and to understand how changes at the molecular level affect development and translate into phenotypic diversity. To achieve this, two distinct approaches have been pursued in recent years. First, comparative studies have revealed strong correlations between the expression patterns of individual regulatory genes during development and differences in morphology^{1–7}. Second, direct genetic analysis has been used to estimate the number and identity of genetic loci that contribute to morphological variation within and between species^{8–10}. Despite their respective successes, the two approaches remain far apart because of their different scales of analysis. Comparative studies have concentrated mainly on slowly evolving traits among high-level taxa, but genetic analyses are only possible among closely related species that produce viable and fertile hybrids.

Our approach to bridging this gap between evolutionary genetics and comparative embryology is to analyse and compare the development of rapidly evolving morphological traits. In many animals, secondary sexual characteristics evolve rapidly^{11,12}, making them good candidates for analysis. One such character in *Drosophila* is the pigmentation of adult abdominal segments. In *D. melanogaster*, abdominal pigmentation is sexually dimorphic. Segments 1 to 6 in females and 1 to 4 in males carry only a posterior stripe of dark pigment. However, segments 5 and 6 (A5 and A6) in males are completely pigmented (Fig. 1a and b), giving the species its name. This pattern is of recent evolutionary origin; in most *Drosophila* species, male-specific pigmentation is absent, so that females and males are pigmented identically (Fig. 1c and d). To understand how this new pattern originated and evolved, we have characterized the regulatory circuit that controls its development, and compared its operation in sexually dimorphic and monomorphic species.

Genetic control of male-specific pigmentation

The non-sex-specific striped pigmentation, which is present in most *Drosophila* species, is controlled by the transcription factor *optomotor-blind* (*omb*)¹³. In *omb* mutants, all pigment is lost from segments A1–A4 in males (Fig. 2a), and from all abdominal segments in females (not shown). However, we find that *omb*[–] males have fully pigmented A6, and mostly pigmented A5 (Fig. 2a). Thus, the ancestral striped pigmentation and the newly evolved male-specific pigmentation are controlled by separate genetic pathways.

The dark pigmentation of A5 and A6 in males is controlled by the homeotic gene *Abdominal-B* (*Abd-B*)^{14,15}. In the pupal abdomen, *Abd-B* is expressed at progressively higher levels in A5, A6 and A7,

but is absent from the more anterior segments (Fig. 3a). Ectopic expression of *Abd-B* in A3 and A4 expands male-specific pigmentation to these segments^{14–17} (Fig. 2b). Conversely, loss of *Abd-B* expression from A5–A7 (ref. 18) eliminates male-specific pigmentation, but has no effect on the non-sex-specific pigment stripes (Fig. 2c). *Abd-B* is the primary activator of male pigmentation, but another homeotic gene, *abdominal-A* (*abd-A*), is expressed in A2–A7 (Fig. 3b) and contributes to the specification of A5 and A6 identities^{14–16}.

The development of sexually dimorphic external characteristics is controlled by the *doublesex* (*dsx*) gene^{19,20}. Alternative splicing of the *dsx* transcript produces a male-specific product in males (*dsxM*), and a female-specific product in females (*dsxF*)^{21,22}. Loss of *dsx* function in females results in the development of male-like pigmentation (Fig. 2d), which can be suppressed by heat-shock *dsxF* transgenes^{23,24}. Male-specific pigmentation is therefore expressed by default, and must be actively repressed by *dsxF*.

Thus, the development of sexually dimorphic pigmentation requires integration of homeotic and sex determination gene inputs. In investigating how this integration is achieved, we discovered a newly evolved genetic circuit that appears to be responsible for the origin of male-specific pigmentation.

bric-a-brac represses male-specific pigmentation

A gene near the left tip of the third chromosome contributes to the

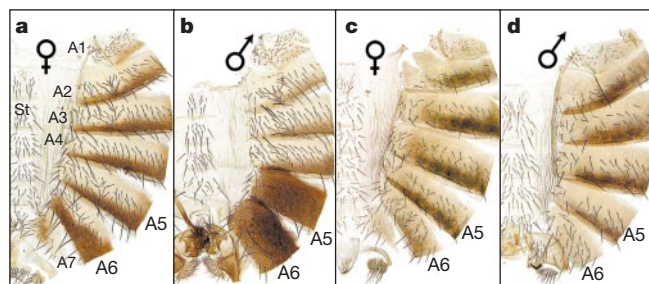


Figure 1 Sexually dimorphic pigmentation is of recent evolutionary origin. *D. melanogaster* female (a) and male (b); and *D. willistoni* female (c) and male (d). Most abdominal segments contain pigmented dorsal and unpigmented ventral cuticular plates (tergite (Ter) and sternite (St), respectively) (see Fig. 6a and b for more detail). In the ancestral condition, pigmentation is similar in males and females (c, d). In the derived condition, abdominal segments 5 and 6 (A5 and A6) are fully pigmented in the male (b), but not in the female (a).

variation in female abdominal pigmentation²⁵. In investigating this genetic region, we found that loss of one copy of the *bric-a-brac* (*bab*) locus results in the development of male-specific pigmentation in females (Fig. 2e), but has no effect on the male abdomen. Ectopic pigmentation in heterozygous *bab* females is suppressed by reducing the dosage of *Abd-B* (Fig. 2f), but is not eliminated by loss of *omb* (not shown). This suggests that *bab*⁺ represses the development of male-specific pigmentation in females by opposing the function of *Abd-B*. The *bab* locus contains two closely related genes, *bab1* and *bab2*, which encode putative transcription factors with multiple roles in development^{26,27}. Ectopic pigmentation in females increases in the order *bab1/+* < *bab1/bab1* < *bab1bab2/+* < *bab1bab2/bab1* (not shown), indicating that both genes are involved

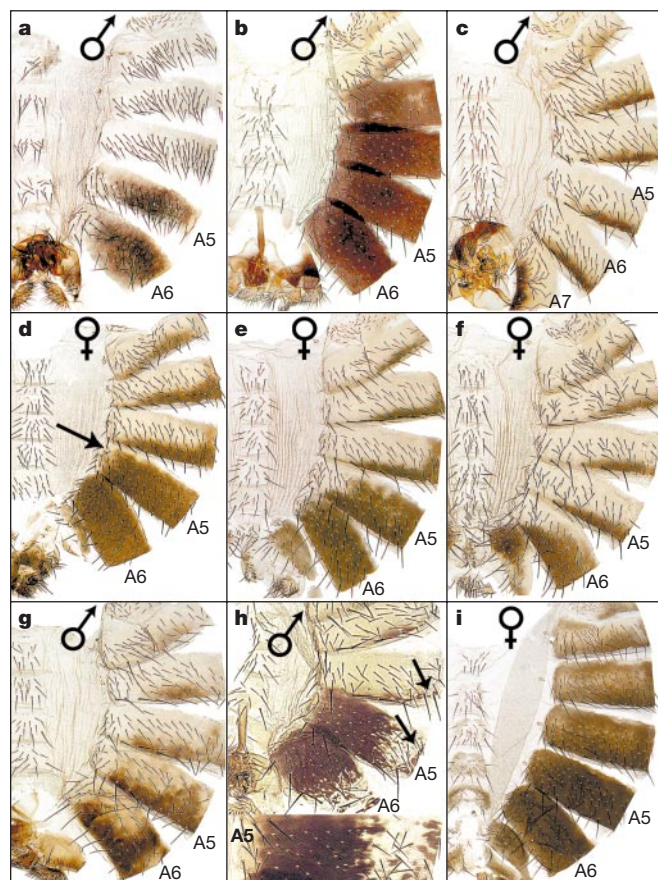


Figure 2 Control of male-specific pigmentation by *Abd-B*, *dsx* and *bab*. **a**, Male hemizygous for the *omb* null mutant *l(1)omb*²⁸² lacks segmental pigment stripes but retains pigmentation in A5 and A6. **b**, In *Abd-B*^{Mcp} *Abd-B*^{Sab} homozygous males, male-specific pigmentation is expanded into A3 and A4. **c**, In *Df(3R)RS4-8/Df(3R)RS1-98* males, loss of *Abd-B* function from A5–A7 eliminates male-specific pigmentation, but has no effect on the non-sex-specific striped pigmentation (see also Fig. 6c). **d**, *dsx*¹/*Df(3R)dsx2D* chromosomal females show male-like pigmentation of A5 and A6, with the exception of a small anterior-lateral margin of A5 (arrow). **e**, *bab*^{Ar07/+} female shows male-specific pigmentation of A5 and A6. **f**, *bab* phenotype is suppressed by reduction of *Abd-B* dosage in *bab*^{Ar07}/*Df(3R)RS4-8* females. **g**, Low-level *bab* expression in *UAS-bab2*⁴⁻⁶⁶ in the absence of a GAL4 driver results in the loss of male-specific pigmentation, but has no effect on the non-sex-specific pigmentation. **h**, High-level *bab* expression in *pnr-GAL4/UAS-bab2*⁴⁻⁵¹ males results in the loss of both male-specific and striped pigmentation (arrows); the latter effect is seen in females as well. An identical phenotype is caused by misexpression of *bab1*. The loss of male-specific pigmentation in this and other *GAL4/UAS-bab* combinations is enhanced by *abd-A* *Abd-B* deficiencies, and partly suppressed by *abd-A* *Abd-B* duplications (not shown). **i**, *bab*^{Ar07}/*bab*^{Ar07} female. Ectopic pigmentation is seen in A2–A7 tergites, A4–A7 sternites, and occasionally in the A1 tergite. An identical pigmentation phenotype is seen in males (not shown).

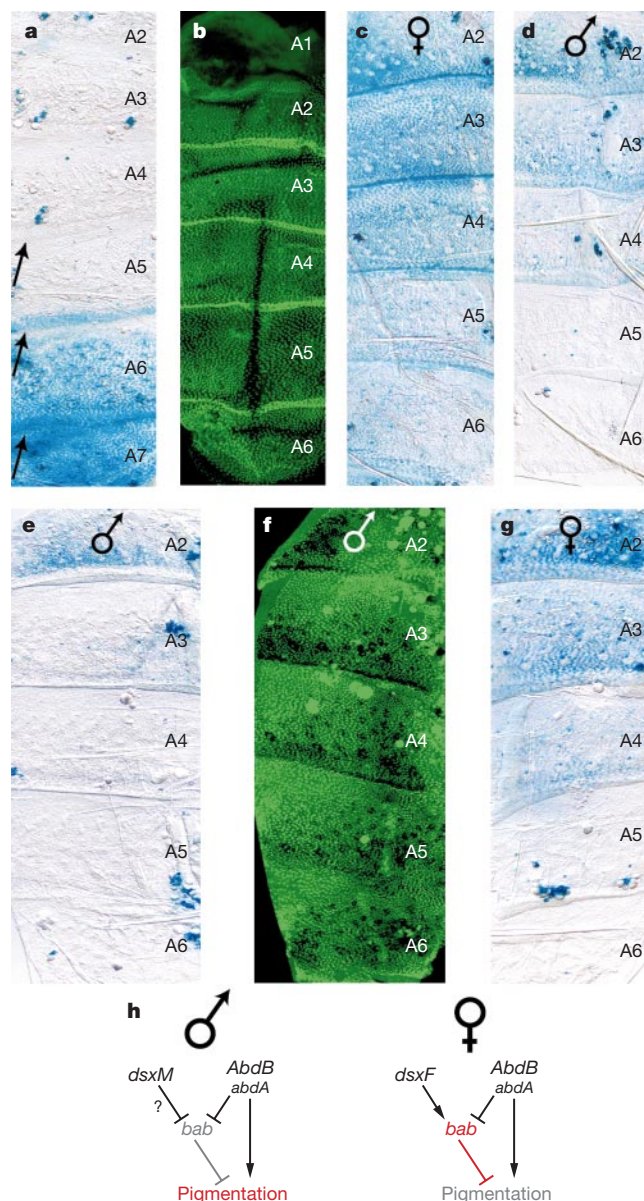


Figure 3 Regulation of *bab* expression by *Abd-B* and *dsx*. All panels show gene expression in the dorsal abdominal epidermis at 40–45 hours after pupariation, shortly before the onset of cuticle differentiation. **a**, *Abd-B-lacZ* female. Segment boundaries are indicated by arrows; *Abd-B* expression is modulated on a parasegmental rather than segmental basis. In A5, expression is only visible in the polyloid bristle cells. **b**, *Abd-A* is expressed in A2–A7. **c**, **d**, *bab1-lacZ* expression in wild-type female and male, respectively. *bab* expression in A5 and A6 is absent in males (**d**) and downregulated in females (**c**). *bab* expression is first seen at 35 hours after pupariation, shortly before the formation of adult epidermis is completed. *bab* expression is modulated parasegmentally, and is not affected by *hh* or *omb* mutations (not shown). *bab* is also expressed in the ventral abdomen, but at a lower level than in the tergites. Little or no expression is seen in A1. **e**, In *Abd-B*^{Mcp} *Abd-B*^{Sab}/*+* males, *bab1-lacZ* is repressed in A3 and A4. **f**, *Bab2* expression is derepressed in A5–A7 in a *Df(3R)RS4-8/Df(3R)RS1-98* male. **g**, *bab1-lacZ* expression in *dsx*¹/*dsx*¹⁸ chromosomal females is similar to that in males (**d**). **h**, Control of sexually dimorphic pigmentation of A5 and A6 by *Abd-B*, *abd-A*, *dsx* and *bab* (see text). It is unclear whether the same circuit operates in A7, as *bab* expression in females is stronger in A7 than in A6. *bab* expression is also present throughout A8–A10, where it is apparently independent of *Abd-B* and *dsx*.

in repressing male pigmentation. For simplicity, we treat the entire locus as one gene, *bab*, unless noted otherwise.

The expression pattern of *bab* at the pupal stage when the adult epidermis develops reflects its sex- and segment-specific function. In females, *bab* expression is strongest in segments A2 and A3, and progressively weaker in A4, A5 and A6 (Fig. 3c). In males, *bab* expression is considerably weaker than in females in all segments. Most strikingly, it is completely absent from A5 and A6 (Fig. 3d). This pattern of *bab* repression correlates with the presence of sex-specific pigmentation in males, and its absence in females.

To test whether *bab*⁺ is sufficient to repress pigmentation, we ectopically expressed the *bab* genes in the pupal abdomen. Low-level expression of *bab*⁺ results in the loss of male-specific pigmentation, but has no other effects on external morphology (Fig. 2g), indicating that differential regulation of *bab* plays a central role in establishing sexual dimorphism. *bab*⁺ can also repress non-sex-specific pigment stripes when expressed at a higher level (Fig. 2h). This suggests that *bab*⁺ acts as a general repressor of pigmentation, but that its effects are overridden by *omb* in the posterior part of each segment. Consistent with this, complete loss of both *bab* genes results in ectopic pigmentation of A2 to A7 in both sexes (Fig. 2i). This phenotype is not caused by expansion of *Abd-B* expression, which appears normal in these mutants (not shown). In *bab* homozygotes, the intensity of pigmentation is higher in the more posterior segments than in those more anterior (Fig. 2i). This suggests that pigmentation does not develop by default in the absence of *bab*, but is actively promoted by *Abd-B* and *abd-A*.

bab integrates regulatory inputs from *Abd-B* and *dsx*

The sexually dimorphic repression of *bab* in the posterior abdomen suggests that *bab* integrates the homeotic and sex determination regulatory inputs. To test this, we examined *bab* expression in *Abd-B* and *dsx* mutant backgrounds. We find that ectopic expression of *Abd-B* in A3 and A4 eliminates *bab* expression from these segments in males (Fig. 3e), and downregulates it in females (not shown). Conversely, *bab* is derepressed in A5–A7 in the mutants that lack *Abd-B* function in these segments (Fig. 3f). Together, these results indicate that *bab* expression in A5 and A6 is normally repressed by *Abd-B*. The slight downregulation of *bab* in A4 (Fig. 3c and d) suggests that it is also weakly repressed by *abd-A*.

In *dsx*[−] intersexes, *bab* is expressed in a male-like pattern (Fig. 3g), suggesting that *dsxF* upregulates *bab* transcription in females. *Abd-B* and *abd-A* expression is identical in males, females and *dsx*[−] intersexes (not shown), indicating that *bab* is regulated independently by homeotic and sex-determination inputs. *dsx*^{Dominant} intersexes, which express both male- and female-specific *dsx* products²⁸, also show male-like expression of *bab* (not shown), indicating that *dsxM* can interfere with *dsxF* function. The two *dsx* isoforms encode transcription factors that bind the same DNA sequence, but have opposite effects on gene expression^{29,30}. *dsx*[−] intersexes differ from males in having a small unpigmented region at the anterior-lateral margin of A5 (refs 19, 23; Fig. 2d), suggesting that *dsxM* may have a slight negative influence on *bab* expression.

Our results suggest that *bab*⁺ regulates sexually dimorphic pigmentation by integrating regulatory inputs from the homeotic genes and the sex determination pathway (Fig. 3h). In this regulatory circuit, *bab*⁺ acts as a general repressor of pigmentation, and *Abd-B* and *abd-A* promote pigmentation in both sexes. In addition, *Abd-B*, and to a lesser extent *abd-A*, repress *bab* transcription. In males, this results in the absence of *bab* from A5 and A6, allowing *Abd-B* and *abd-A* to promote pigmentation in these segments. However, in females, *dsxF* prevents *bab* transcription from being completely repressed by the homeotic genes. As a result, *bab* is present in A5 and A6 in females, where it blocks the ability of *Abd-B* and *abd-A* to promote pigmentation. In A2–A4, *abd-A* alone is not sufficient either to repress *bab* or to overcome its inhibitory effect on pigmentation; thus, only the *omb*-dependent striped pigmentation

is generated. Because *Abd-B*, *abd-A* and *dsx* encode transcription factors, they may regulate *bab* expression directly.

Evolution of sexually dimorphic pigmentation

The central role of *bab* as an integrator of homeotic and sex-determination gene inputs suggests that changes in *bab* regulation may have been responsible for the evolution of sexually dimorphic pigmentation. In the subgenus *Sophophora*, male-specific pigmentation is present only in the *melanogaster* species group. Within this group, sexual dimorphism is seen in all species of the *melanogaster* subgroup and the closely related oriental subgroups, whereas the *ananassae* and *montium* subgroups contain both sexually dimorphic and sexually monomorphic species.

We find that in species with male-specific pigmentation of A5 and A6, *bab* expression is absent or strongly downregulated in these segments in males, but not in females (Fig. 4a and b; Fig. 5). Moreover, in the sexually monomorphic species outside the *melanogaster* species group, *bab* expression is identical in both sexes and in all segments from A2 to A7 (Fig. 4c and d; Fig. 5). This correlation suggests that changes in the regulation of *bab* by *Abd-B* and *dsx* played an important role in the origin of sexually dimorphic pigmentation.

Surprisingly, in some species of the *montium* subgroup, such as *D. kikkawai*, *bab* expression in A5 and A6 is downregulated in males despite the absence of sex-specific pigmentation (Fig. 5). This suggests that the function of *bab*⁺ may not be limited to the control of pigmentation. To test whether *bab*⁺ regulates other morphological characters in abdominal segments, we analysed the phenotypes produced by gain and loss of *bab* function in *D. melanogaster*.

Homeotic function of *bab*

In *D. melanogaster*, A5, A6 and A7 differ from the more anterior segments not only in pigmentation, but also in the size and shape of tergites and sternites and in the distribution of bristles and trichomes (Fig. 6a and b). These characteristics, which are especially pronounced in the male, are under the control of *Abd-B*, and the differences among segments are thought to be specified by the different levels of *Abd-B* expression^{14–18}. Ectopic expression of *Abd-B* induces posterior characters in anterior segments, while the loss of *Abd-B* function from A5–A7 transforms these segments to a more anterior identity (Fig. 6c).

We find that *bab*⁺ regulates segment shape and bristle and trichome patterns in a manner reciprocal to *Abd-B*. Loss of *bab*⁺ function in females enhances posterior characteristics in A6, A7 and A8 (Fig. 6d). No phenotype is seen in males, consistent with the absence of *bab* expression in posterior segments. Conversely, ectopic expression of *bab* transforms A6 and A7 to a more anterior identity in both males (Fig. 6e) and females (see Supplementary Information). These observations suggest that *bab*⁺ acts as an antagonist of

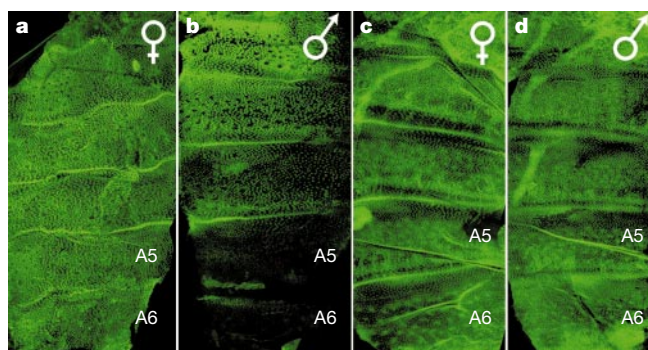


Figure 4 Repression of *bab* correlates with the presence of male-specific pigmentation. *Bab2* expression is modulated in the sexually dimorphic species *D. biarmipes* (a, female; b, male), but not in the sexually monomorphic *D. willistoni* (c, female; d, male).

Abd-B homeotic function, and that posterior abdominal characters are determined by the balance between *Abd-B* and *bab* activities.

This model predicts that evolutionary changes in *bab* regulation should result in morphological transformation of *Abd-B*-expressing segments. Indeed, we find that the entire suite of characteristics that distinguishes A5 and A6 from the more anterior segments in *D. melanogaster* is of recent evolutionary origin. In *D. willistoni*, *bab* is expressed strongly in A5 and A6 in males (Fig. 4d), whereas *Abd-B* is expressed in the same pattern as in *D. melanogaster* (not shown). As predicted, A5 and A6 are almost identical to the more anterior, non-*Abd-B*-expressing segments in the males of this species (Fig. 6f). In contrast, the *melanogaster* species group shows great diversity of bristle and trichome patterns in posterior abdominal segments (Table 1). The two main lineages within this group show different patterns of evolution. In the clade composed of the *melanogaster* and oriental subgroups, male-specific pigmentation and bristle and trichome patterns have evolved in a concerted fashion (Table 1, group A). However, in the *ananassae* + *montium*

lineage, these characteristics vary independently of each other, and sexually dimorphic bristle and trichome patterns are sometimes observed in species that do not show visible modulation of *bab* expression (Table 1, group B, compare with Fig. 5). This suggests that evolutionary changes have occurred not only in *bab* regulation, but also in the target genes of *bab* and in other genes regulated by *Abd-B* and *dsx*. We also note that suppression of A7 development in males has occurred earlier in evolution than visible modulation of *bab* expression, despite the ability of *bab* to override this suppression (Fig. 6e).

Sexually dimorphic pigmentation affects mating preferences

The rapid evolution of sexually dimorphic pigmentation and segment morphology may have been driven by sexual selection. We therefore tested whether male-specific pigmentation confers a competitive advantage in *D. melanogaster* males. Surprisingly, we find that UAS-*bab2*⁴⁻⁶⁶ males, which lack male-specific pigmentation but are otherwise normal (Fig. 2g), enjoy the same mating

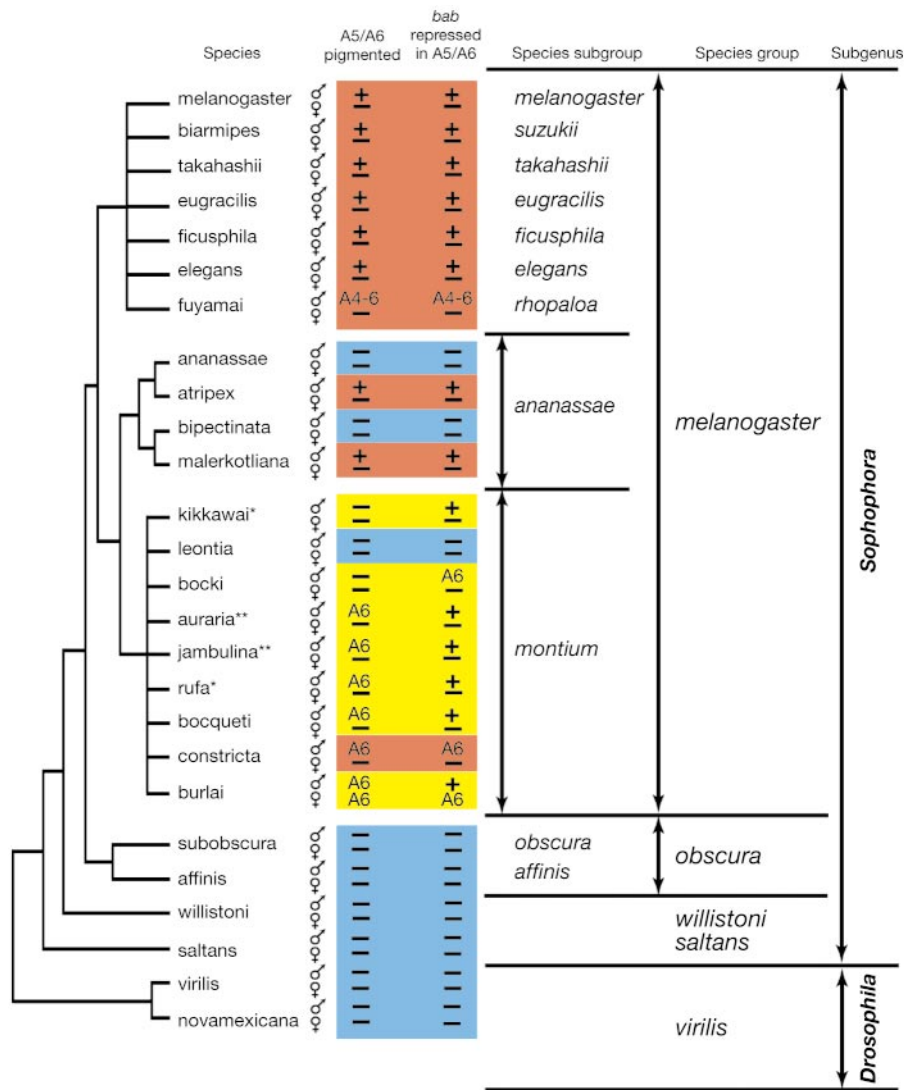


Figure 5 Modulation of *bab* expression correlates with sexual dimorphism. Phylogenetic hypothesis is based on information from various sources⁴³⁻⁴⁶. The divergence time between *melanogaster* and *obscura* species groups has been estimated at 25 million years⁴⁷. A6 pigmentation is variable in the females of *D. kikkawai* and *D. rufa* (asterisk), and in both females and males of *D. auraria* and *D. jambulina* (double asterisk). Bab2 is repressed in A5 and A6 in the males of sexually dimorphic species (red), but not in the sexually monomorphic species (blue). Exceptions to this rule are found among species of

the *montium* subgroup, in which Bab2 is downregulated in the absence of male-specific pigmentation (yellow). Expression was only analysed in the dorsal abdomen. We cannot rule out that species listed as having unmodulated *bab* expression do in fact have subtle modulation that is below the resolution of our methods. For instance, a twofold reduction in the dosage of *bab*, which causes a dramatic pigmentation phenotype in the females of *D. melanogaster* (Fig. 2e), is not detectable by antibody staining. For the same technical reason, quantitative differences between sexes and species were not studied.

success as wild-type males (27 versus 32 matings; $\chi^2 = 0.42$; $P > 0.05$). Thus, although male pigmentation may have been important in the past, it appears to have little or no effect on female mating preferences in extant *D. melanogaster*.

However, we find that *D. melanogaster* males discriminate strongly against heterozygous *bab* females, which have ectopic male-specific pigmentation but are otherwise normal (Fig. 2e), compared with females with lightly pigmented A5 and A6 (23 versus 105 matings; $\chi^2 = 52.53$; $P < 0.001$). Importantly, white mutant males, which are visually impaired, mate equally with *bab*/+ and lightly pigmented females (34 versus 39 matings; $\chi^2 = 0.34$; $P > 0.05$), suggesting that discrimination against *bab* heterozygous females is due to their pigmentation. These results suggest that female pigmentation is important in determining their attractiveness to males, and that the absence of male-specific pigmentation in females may be maintained by sexual selection.

A model of evolution of a genetic circuit under sexual selection

Our findings indicate that changes in *bab* regulation have played an important part in the evolution of abdominal segment morphology (Fig. 6g). The presence of *bab* expression in all *Drosophila* species examined suggests that its roles in antagonizing the homeotic function of *Abd-B* and repressing pigmentation are ancestral.

However, in the ancestral condition, *bab* expression was independent of *Abd-B* and *dsx*, resulting in sexually monomorphic pigmentation and segment morphology. In the *melanogaster* species group, *bab* evolved to be under the control of *Abd-B* and *dsx*. This eliminated *bab* from *Abd-B*-expressing segments in the male and resulted in a major transformation of male segment morphology. Subsequent diversification of pigmentation, bristle and trichome patterns was probably driven both by the fine-tuning of *bab* regulation and by changes in the downstream targets of *bab* and *Abd-B*.

Two features of this genetic circuit make it highly plastic and evolvable. First, the adult phenotype is sensitive to quantitative changes in *bab* expression. Second, the level of *bab* expression is determined by the balance between *Abd-B* and *dsxF* inputs. If *bab* is regulated directly by *Abd-B* and *dsx*, then the evolution of sexually dimorphic pigmentation and segment morphology may ultimately be traced to the acquisition and modification of binding sites for the *Abd-B* and *Dsx* proteins in the *cis*-regulatory region of *bab*. Thus, even a subtle molecular change could be expressed phenotypically and become subject to selection.

This evolutionary model is further supported by the presence of intraspecific genetic variation in sexually dimorphic pigmentation in many extant species^{31,32}. In at least one case, there is strong

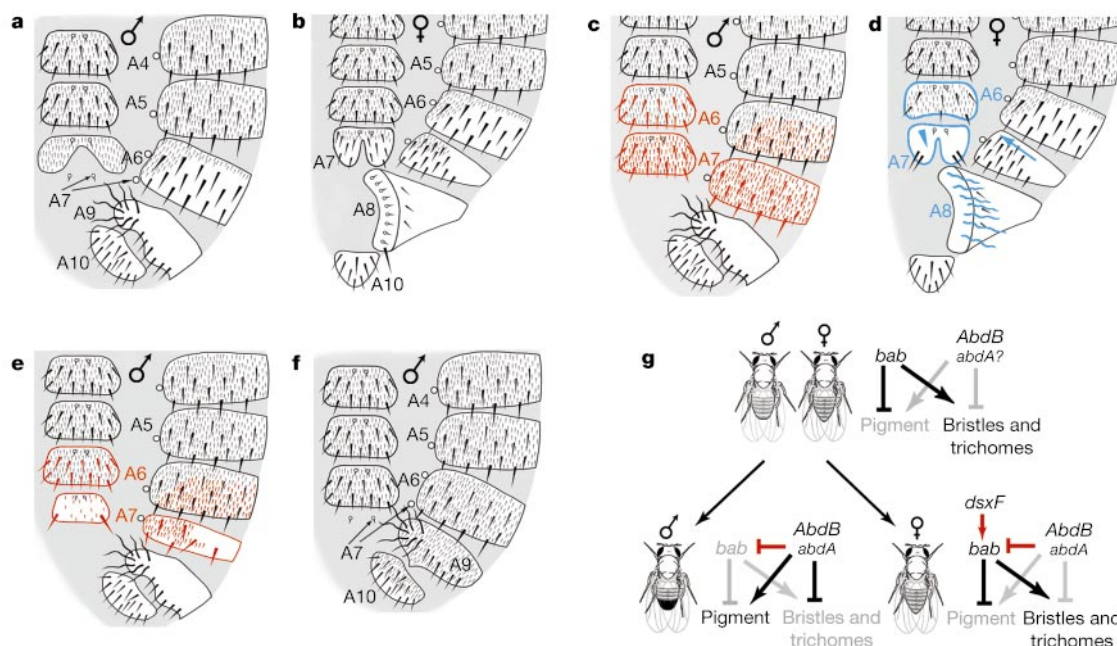


Figure 6 The homeotic function of *bab*. Photomicrographs of the cuticles are available in the Supplementary Information. **a**, Wild-type male of *D. melanogaster*. A7 is greatly reduced and lacks both tergite and sternite. In addition, the A6 sternite lacks bristles and has a unique horseshoe-like shape, and most of the A6 tergite is devoid of the small hairs (trichomes) that cover the more anterior segments. Male genitalia develop from A9, whereas A8 development is repressed. **b**, In a wild-type female, A7 sternite and A7 and A6 tergites are only partly covered by trichomes; the A7 sternite also has a distinctive shape. Female genitalia develop from A8, whereas A9 development is repressed. **c**, In *Df(3R)RS4–8/Df(3R)RS1–98* males, A5–A7 are transformed towards A4 identity¹⁸ (indicated in red). The presence of sternite and tergite in A7 is restored, A6 and A7 tergites are completely covered by trichomes, and A6 and A7 sternites are rectangular and carry bristles. **d**, In *bab^{Ar07}/bab^{Ar07}* females, expression of posterior morphological characters is increased (blue). Trichomes are absent from the A7 tergite and sternite, and the number of bristles on A6 and A7 sternites is reduced. The A6 sternite also acquires a shape reminiscent of the A6 in the male, or A7 in the female. A weaker phenotype is seen in *bab^{Ar07}/+* (not shown). The thorn bristles normally present on the vaginal plates of A8 (**b**) are replaced by long, wavy bristles typical of the male genital arch (A9). *bab⁻* females are nevertheless distinguishable from males in the morphology of A6 and particularly A7,

indicating that *dsxM* and/or *dsxF* regulate additional target genes in the abdomen. **e**, Ubiquitous *bab* expression in *C765-GAL4/UAS-bab^{24–51}* males transforms A6 and A7 to a more anterior identity (red). The presence of tergite and sternite in A7 is restored, A6 sternite acquires a rectangular shape and a number of bristles, and the entire A6 tergite is covered by trichomes. A similar phenotype is caused by misexpression of *bab1* (not shown). The phenotype is enhanced by *abd-A Abd-B* deficiencies, and partly suppressed by *abd-A Abd-B* duplications (not shown). Ectopic *bab* has no effect in the genitalia and analia of either sex (segments A8–A10). The homeotic function of *bab* is apparently limited to the adult stage, because ectopic *bab* expression has no effect in the larval cuticle (not shown). **f**, In *D. willistoni* males, the A6 sternite is rectangular and carries numerous bristles, and the A6 tergite is covered by dense trichomes. We note the similarity to **c** and **e**. **g**, The role of *bab* in the evolution of sexually dimorphic segment morphology. We suggest that sexual dimorphism evolved through the acquisition of two new regulatory interactions, shown in red. The number of genetic steps involved in this transition is unclear. In the *ananassae* and *montium* subgroups, *bab* regulation may differ strongly between closely related species (Fig. 5), indicating that the new genetic circuit retained considerable plasticity.

Table 1 Male-specific characters in two lineages of *melanogaster* species group*

Species	A5/A6 pigmentation (A6 only in <i>montium</i> subgroup)	Loss of A6 tergite trichomes	Loss of A6 sternite bristles
A			
<i>D. melanogaster</i>	+	+	+
<i>D. biarmipes</i>	+	+	+
<i>D. takahashii</i>	+	+	+
<i>D. eugracilis</i>	+	+	+
<i>D. ficusphila</i>	+	+	+
<i>D. elegans</i>	+	—	+
<i>D. fuyamai</i>	+	+	+
B			
<i>D. ananassae</i>	—	—	+
<i>D. atripex</i>	+	—	+
<i>D. leontia</i>	—	—	+
<i>D. bocki</i>	—	+	+
<i>D. auraria</i>	+/-†	+	—
<i>D. rufa</i>	+	+	—
<i>D. burlai</i>	+	+	+
<i>D. nikananu</i>	—	+	+

* Photomicrographs of some species are available in the Supplementary Information.

† A6 pigmentation is variable in *D. auraria* males.

evidence that allelic differences at the *bab* locus contribute to this variation. Females found in natural populations of *D. melanogaster* vary widely in the extent of A6 pigmentation, ranging from near-zero to 100% (ref. 25). The locus with the largest effect on this variation has been mapped²⁵—to the exact position of *bab*. These observations suggest that sexually dimorphic pigmentation evolved through fixation of intraspecific genetic variants at the *bab* locus.

Fixation of new *bab* alleles was probably driven initially by “runaway” sexual selection. In this case, a slight female preference for a weakly pronounced male character would initiate a positive feedback loop that would rapidly increase both the expression of the male character and the female preference for it³³. This self-reinforcing mechanism can drive rapid character divergence and create new species through sexual isolation^{34–37}. Male-specific pigmentation could evolve by this mechanism, with increasingly discriminating females selecting for increasingly dark males. However, once fixed, sexual characteristics can lose their significance as they are overtaken by newly evolving signals and as females become habituated and “resistant” to old characters³⁸. This may explain our finding that male pigmentation has no effect on mating success in extant *D. melanogaster*.

Whereas the runaway model explains the evolution of male sexual characters, it does not account for the absence of these characters in females, that is, sexual dimorphism. However, sexual dimorphism can be produced effectively by counter-selection against male-specific traits in females³³. Consistent with this, we find that *D. melanogaster* males discriminate against females that have male-like pigmentation. In most *Drosophila* species, including *D. melanogaster*, males seek out females at feeding sites and attempt to court as many as possible. Courting other males is not only disadvantageous in competition for females, but may also carry a direct cost³⁹. Thus, males are probably selected for an ability to avoid courting other males, and pigmentation may be used to identify females at a distance.

The evolution of *bab* regulation offers a tractable model of how selection creates new morphological characters through changes in DNA sequence. Analysis of the *cis*-regulatory elements of *bab* in sexually dimorphic and monomorphic species will help to clarify the molecular basis of morphological divergence between these taxa. □

Methods

Anterior is upward in all figures. Adult abdomens were cut along the dorsal midline and mounted flat so that ventral cuticle is on the left and dorsal cuticle is on the right. Pupal dissections, X-Gal staining⁴⁰ and antibody staining⁴ were performed as described.

Drosophila strains and reagents

Abd-B^{Mcp} Abd-B^{Sab} is a double gain-of-function mutation that results in ectopic *Abd-B* expression in A3 and A4 (ref. 15). *Df(3R)RS4–8* removes the *cis*-regulatory elements responsible for *Abd-B* expression in A5–A7; in *Df(3R)RS4–8/Df(3R)RS1–98* heterozygotes, these segments lack *Abd-B* function¹⁸. *Abd-B-lacZ* enhancer trap HCJ199 (ref. 41) reproduces the pattern of *Abd-B* protein expression (not shown). *Abd-A* protein expression was detected by monoclonal antibody Dmabd-A.1 (ref. 42).

bab^{Ar07} (F. Laski, unpublished work) is a small deletion that inactivates both *bab* genes. *UAS-bab^{24–66}* (D. Godt, unpublished work) is inserted into the 5' regulatory region of the ecdyson-inducible gene *ftz-F1* (data not shown). The phenotype produced by this line in the absence of a GAL4 driver is presumably caused by low-level expression of *bab2* driven by *ftz-F1* regulatory elements during metamorphosis. In the presence of GAL4, this line generates phenotypes similar to other *UAS-bab* insertions, including *UAS-bab^{4–51}*. *pnr-GAL4* drives gene expression in a stripe along the dorsal midline in all abdominal segments. *bab1-lacZ* enhancer trap²⁶ reproduces the expression of both *bab* genes. *Bab1* and *Bab2* proteins, detected by specific antibodies (F. Laski, unpublished work), are expressed in similar patterns in all species examined (*melanogaster*, *virilis*, *willistoni*, *ananassae* and *eugracilis*) (data not shown).

Mating choice experiments

Virgin males and females, 3–5 days old, were placed together in a 95 × 24mm glass vial, and matings were observed visually. For male-choice experiments, a single male was combined with two females of different genotypes. For female-choice experiments, a single female was combined with two males of different genotypes. Only matings that occurred in the first 30 min were counted; the number of unmated groups was also recorded. The *bab* genotypes used were *bab^{Ar07}/+*, *bab^{Ar07}/TM3*, *bab^{FC}/bab-lacZ* and *Df(3L)Ar12–1/+*. Strains in which females have lightly pigmented A6 were isolated from various laboratory stocks, and were not standard wild-type strains. For some experiments, *bab^{Ar07}* was introgressed into a ‘light’ strain for 3–4 generations. Oregon-R strain was used as a source of wild-type males. For the most part, identical results were obtained with different *bab* alleles and wild-type strains, as well as following the introgression of *bab^{Ar07}* into a lightly pigmented strain, suggesting that mating discrimination is not due to differences in genetic backgrounds.

In experiments with one ‘light’ strain, wild-type males mated equally with *bab^{Ar07}/+* and lightly pigmented females (18 versus 22 matings; $\chi^2 = 0.4$; $P > 0.05$). In these experiments, males courted both types of females repeatedly; however, courtship was quickly aborted in most cases, and nearly 50% of males did not mate within 1 hour. The reasons for this behaviour are unclear.

Received 30 June; accepted 10 October 2000.

- Warren, R., Nagy, L., Selegue, J., Gates, J. & Carroll, S. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature* **372**, 458–461 (1994).
- Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. *Hox* genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).
- Averof, M. & Akam, M. *Hox* genes and the diversification of insect-crustacean body plans. *Nature* **376**, 420–423 (1995).
- Brakefield, P. M. *et al.* Development, plasticity and evolution of butterfly eyespot patterns. *Nature* **384**, 236–242 (1996).
- Averof, M. & Patel, N. H. Crustacean appendage evolution associated with changes in *Hox* gene expression. *Nature* **388**, 682–686 (1997).
- Stern, D. L. A role of *Ultrabithorax* in morphological differences between *Drosophila* species. *Nature* **396**, 463–466 (1998).
- Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo/shaven-baby*. *Proc. Natl Acad. Sci. USA* **97**, 4530–4534 (2000).

8. True, J. R., Stam, L. F., Zeng, Z. -B. & Laurie, C. C. Quantitative genetic analysis of divergence in male secondary sexual traits between *Drosophila simulans* and *Drosophila mauritiana*. *Evolution* **51**, 816–832 (1997).
9. Gurganus, M. C., Nuzhdin, S. V., Leips, J. W. & MacKay, T. F. High-resolution mapping of quantitative trait loci for sternopleural bristle number in *Drosophila melanogaster*. *Genetics* **152**, 1535–1604 (1999).
10. Lyman, R. F., Lai, C. & MacKay, T. F. Linkage disequilibrium mapping of molecular polymorphisms at the *scabrous* locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genet. Res.* **74**, 303–311 (1999).
11. Anderson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
12. Eberhard, W. G. Copulatory courtship and cryptic female choice in insects. *Biol. Rev.* **66**, 1–31 (1991).
13. Kopp, A. & Duncan, I. Control of cell fate and polarity in the adult abdominal segments of *Drosophila* by *optomotor-blind*. *Development* **124**, 3703–3714 (1997).
14. Sanchez-Herrero, E., Vernos, I., Marco, R. & Morata, G. Genetic organization of the *Drosophila* bithorax complex. *Nature* **313**, 108–113 (1985).
15. Duncan, I. M. The bithorax complex. *Annu. Rev. Genet.* **21**, 285–319 (1987).
16. Celniker, S. E., Sharma, S., Keelan, D. J. & Lewis, E. B. The molecular genetics of the bithorax complex of *Drosophila*: cis-regulation in the *Abdominal-B* domain. *EMBO J.* **9**, 4277–4286 (1990).
17. Crosby, M. A., Lundquist, E. A., Tautvydas, R. M. & Johnson, J. The 3' regulatory region of the *Abdominal-B* gene: genetic analysis supports a model of reiterated and interchangeable regulatory elements. *Genetics* **134**, 809–824 (1993).
18. Hopmann, R., Duncan, D. & Duncan, I. Transvection in the *iab-5,6,7* region of the bithorax complex of *Drosophila*: homology independent interactions in *trans*. *Genetics* **139**, 815–833 (1994).
19. Baker, B. S. & Ridge, K. A. Sex and the single cell. I. On the action of major loci affecting cell determination in *Drosophila melanogaster*. *Genetics* **94**, 383–423 (1980).
20. Cline, T. W. & Meyer, B. J. Vive al difference: males vs. females in flies vs. worms. *Annu. Rev. Genet.* **30**, 637–702 (1996).
21. McKeown, M. Sex differentiation: the role of alternative splicing. *Curr. Opin. Genet. Dev.* **2**, 299–303 (1992).
22. Burtis, K. C. The regulation of sex determination and sexually dimorphic differentiation in *Drosophila*. *Curr. Opin. Cell Biol.* **5**, 1006–1014 (1993).
23. Jursnich, V. A. & Burtis, K. C. A positive role in differentiation for the male doublesex protein of *Drosophila*. *Dev. Biol.* **155**, 235–249 (1993).
24. Waterbury, J. A., Jackson, L. L. & Schedl, P. Analysis of the doublesex female protein in *Drosophila melanogaster*: Role in sexual differentiation and behavior and dependence on intersex. *Genetics* **152**, 1653–1667 (1999).
25. Robertson, A., Briscoe, D. A. & Louw, J. H. Variation in abdomen pigmentation in *Drosophila melanogaster* females. *Genetica* **47**, 73–76 (1977).
26. Godt, D., Couderc, J. -L., Cramton, S. E. & Laski, F. Pattern formation in the limbs of *Drosophila*: *bric a bric* is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. *Development* **119**, 799–812 (1993).
27. Pointud, J. C. Fly base personal communications report (<http://flybase.bio.indiana.edu/bin/fbpcq.html?FBrf0104845>) (1998).
28. Nagoshi, R. N. & Baker, B. S. Regulation of sex-specific RNA splicing at the *Drosophila doublesex* gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes Dev.* **4**, 89–97 (1990).
29. Burtis, K. C., Coschigano, K. T., Baker, B. S. & Wensink, P. C. The Doublesex proteins of *Drosophila melanogaster* bind directly to a sex-specific yolk protein gene enhancer. *EMBO J.* **10**, 2577–2582 (1991).
30. Coschigano, K. T. & Wensink, P. C. Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes Dev.* **7**, 42–54 (1993).
31. Bock, I. R. Taxonomy of the *Drosophila bipectinata* species complex. *Univ. Texas Publ.* **7103**, 273–280 (1971).
32. Ohnishi, S. & Watanabe, T. K. Genetic analysis of color dimorphism in the *Drosophila montium* subgroup. *Jpn J. Genet.* **60**, 355–358 (1985).
33. Fisher, R. A. *The Genetical Theory Of Natural Selection* (Dover, New York, 1958).
34. Carson, H. L. in *Ecological Genetics: The Interface* (ed. Brussard, P. F.) 93–107 (Springer, New York, 1978).
35. Lande, R. Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
36. Kirkpatrick, M. Sexual selection and the evolution of female choice. *Evolution* **36**, 1–12 (1982).
37. Carson, H. L. in *Evolutionary Genetics* (eds Singh, R. S. & Krimbas, C. B.) 495–512 (Cambridge Univ. Press, Cambridge, 1999).
38. Holland, B. & Rice, W. R. Chase-away sexual selection: antagonistic seduction versus resistance. *Evolution* **52**, 1–7 (1998).
39. Cordts, R. & Partridge, L. Courtship reduces longevity of male *Drosophila melanogaster*. *Anim. Behav.* **52**, 269–278 (1996).
40. Kopp, A. & Duncan, I. Technical tips for analyzing gene expression in the pupal abdomen. *Dros. Info. Serv.* (in the press).
41. Bender, W. & Hudson, A. P. Element homing to the *Drosophila* bithorax complex. *Development* **127**, 3981–3992 (2000).
42. Kellerman, K. A. Mutations affecting the stability of the *fushi tarazu* protein of *Drosophila*. *Genes Dev.* **4**, 1936–1950 (1990).
43. Bock, I. R. & Wheeler, M. R. The *Drosophila melanogaster* species group. *Univ. Texas Publ.* **7213**, 1–102 (1972).
44. Bock, I. R. Current status of the *Drosophila melanogaster* species-group (Diptera). *Syst. Entomol.* **5**, 341–356 (1980).
45. Pelandakis, M. & Solignac, M. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. *J. Mol. Evol.* **37**, 525–543 (1993).
46. Toda, M. J. *Drosophilidae* (Diptera) in Myanmar (Burma). VII. The *Drosophila melanogaster* species-group, excepting the *D. montium* species-subgroup. *Orient. Insects* **25**, 69–94 (1991).
47. Russo, C. A. M., Takezaki, N. & Nei, M. Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**, 391–404 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank F. Laski for *bab* stocks and antibodies; D. Godt for UAS-*bab* lines; W. Bender, S. Celniker and E. Sanchez-Herrero for the *Abd-B* enhancer trap and antibodies; J. David, Y. Fuyama, M. Kimura, J. Roote and the Bowling Green stock centre for various *Drosophila* species; L. Olds for the artwork; B. Holland, S. Nuzhdin, M. Servadio and J. True and T. Wittkopp for discussions. A.K. is a Howard Hughes Medical Institute fellow of the Life Sciences Research Foundation; I.D. is supported by an NIH grant; S.B.C. is an investigator at the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.B.C. (e-mail: sbcarrol@facstaff.wisc.edu).

Evidence against a redshift $z > 6$ for the galaxy STIS123627+621755

Daniel Stern*, Peter Eisenhardt*, Hyron Spinrad†, Steve Dawson†, Wil van Breugel‡, Arjun Dey§, Wim de Vries‡ & S. A. Stanford‡||

* Jet Propulsion Laboratory, California Institute of Technology, Mail Stop 169-327, Pasadena, California 91109, USA

† Department of Astronomy, University of California at Berkeley, Berkeley, California 94720, USA

‡ Institute of Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, L-413, Livermore, California 94550, USA

§ KPNO/NOAO, 850 N. Cherry Avenue, PO Box 26732, Tucson, Arizona 85726, USA

|| Physics Department, University of California at Davis, Davis, California 95616, USA

The identification of galaxies at extreme distances provides the most direct information about the earliest phases of galaxy formation. But at redshifts $z > 5$ even the most luminous galaxies appear faint; the interpretation of low signal-to-noise ratio data is difficult and misidentifications do occur. Here we report optical and near-infrared observations of the source STIS123627+621755, which was previously suggested to be at a redshift of 6.68 (ref. 1). At that redshift, and with the reported¹ spectral energy distribution, the galaxy should be essentially invisible at wavelengths less than 9,300 Å, because the intervening intergalactic medium absorbs almost all light energetic enough to ionize neutral hydrogen—that is, with wavelengths less than the redshifted Lyman limit of $\lambda = (1+z) \times 912$ Å. At near-infrared wavelengths, however, the galaxy should be relatively bright. Here we report a detection of the galaxy at 6,700 Å and a non-detection at a wavelength of 1.2 µm, contrary to expectations for $z \approx 6.68$. The data conservatively require that STIS123627+621755 has a redshift $z < 6$.

During UT 1997 December 23–26, while the Wide Field Planetary Camera 2 obtained second epoch images of the Hubble Deep Field² (HDF), the Space Telescope Imaging Spectrograph (STIS) obtained 4.5 h of filterless imaging (central wavelength $\lambda_c = 5,850$ Å; full-width at half-maximum, FWHM = 4,410 Å) and 13.5 h of slitless spectroscopy of a $52'' \times 52''$ field 4.7 arcmin north-northwest of the HDF. The STIS field is sufficiently separated from the HDF that it lies outside the initial HDF flanking field imaging surveys; that is no supporting photometry is provided by the U to K imaging of refs 2–4. Chen *et al.*¹ have analysed the deep STIS data to identify distant protogalaxies and report that one faint source, STIS123627+621755 (“galaxy A” in ref. 1, has a probable redshift of $z = 6.68 \pm 0.005$,

which would make it the most distant object identified spectroscopically. The source is marginally resolved and has $AB(\text{clear}) = 27.7 \pm 0.1$. The redshift is based upon an emission line at $9,337 \pm 6$ Å with a flux density of $(2.6 \pm 0.5) \times 10^{-17}$ ergs cm⁻² s⁻¹, identified as Lyman α , and an extremely strong continuum break at about 9,300 Å, identified as absorption from the Ly α forest.

For $z = 6.68$, the flux detected in the filterless image must derive almost exclusively from wavelengths longward of the Ly α forest, implying STIS123627+621755 has an apparent flux density $f_\nu \approx 1$ µJy (AB magnitude ≈ 24) longward of 9,300 Å, and should be essentially undetectable at optical wavelengths shortward of 9,300 Å. Assuming this source is a star-forming galaxy as suggested by ref. 1, the inferred rest-frame 1,700 Å absolute magnitude is $M_{AB}(1,700 \text{ Å}) - 5 \log h_{50} = -22.3$ (assuming a flat spectrum source longward of Ly α , $f_\nu \propto \nu^0$, and the current value of the Hubble constant $H_0 = 50 h_{50}$ kms⁻¹ Mpc⁻¹, the cosmological density parameter $\Omega = 1$, and the cosmological constant $\Lambda = 0$). This is more luminous than all but a handful of the more than 700 galaxies at $z > 2$ identified by Steidel *et al.*⁵ over a field-of-view more than a thousand times larger. If STIS123627+621755 is a quasar at $z = 6.68$, the inferred blue luminosity is less atypical of the population with $M_B \approx -22.9$ (assuming $f_\nu \propto \nu^{-0.5}$). However, the probability of identifying a faint ($M_B \leq 22.5$) quasar at $z \geq 6$ in the STIS field-of-view is about 3×10^{-4} (see ref. 6).

These calculations show that STIS123627+621755 is potentially a very unusual source, providing us with a uniquely luminous probe of the very early Universe. We therefore targeted it for deep studies with the Keck telescopes (Fig. 1). We do not detect STIS123627+621755 in our J-band image, which implies that $J_{AB} > 25.3$ (3σ limit in a $2.0''$ diameter aperture). We report a 6σ detection of STIS123627+621755 in our R-band image: $R_{AB} = 26.8 \pm 0.1$ measured in a $0.9''$ -diameter aperture to avoid the neighbouring galaxy to the east with aperture corrections applied assuming the source is unresolved. Between 1998 and 2000 we have also attempted slitmask spectroscopy of STIS123627+621755 numerous times, never robustly detecting the source. As an example of one observation with good conditions and confident astrometry, on UT 2000 April 06 we observed STIS123627+621755 for 2 h with the Low Resolution Imaging Spectrometer⁸ and the 150 lines mm⁻¹ grating. Our 3σ limit to spatially ($1.0''$ FWHM) and spectrally ($\sigma \approx 230$ km s⁻¹) unresolved line emission at 9,350 Å is approximately 2×10^{-17} erg cm⁻² s⁻¹, marginally at odds with the results of ref. 1. We note that the fringing and telluric OH emission which affect ground-based charge-coupled device (CCD) observations at long wavelength make this limit wavelength-dependent.

Our results strongly imply that STIS123627+621755 cannot be at $z > 6$. In Fig. 2 we present the photometric and spectrophotometric data for STIS123627+621755 overplotted with a model star-forming galaxy at $z = 6.68$. Our 1.2-µm non-detection is severely at odds

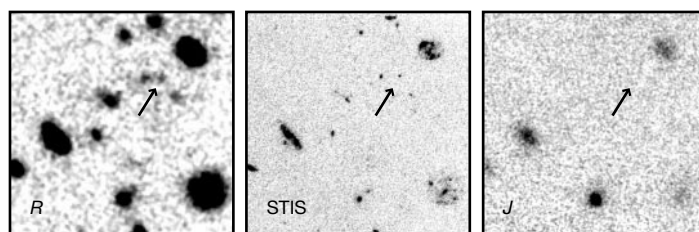


Figure 1 Imaging of STIS123627+621755 showing that the galaxy is detected at R, below the Lyman limit if $z = 6.68$. Fields are $20.0''$ on a side, with north at the top and east to the left. The 70-minute interference R-band ($\lambda_c = 6,700$ Å; $\Delta\lambda = 1,400$ Å) image was obtained on UT 2000 May 05 with the Echelle Spectrograph and Imager on the Keck II telescope in photometric conditions and $0.85''$ seeing. The image has been smoothed by a circular gaussian of width 1 pixel ($0.153''$). The 64-minute J-band ($\lambda_c = 1.25$ µm; $\Delta\lambda = 0.3$ µm) image was obtained on UT 2000 January 30 with the

Near Infrared Camera⁷ on the Keck I telescope in photometric conditions and $0.9''$ seeing. A $3.0''$ -long arrow indicates the location of STIS123627+621755 ($\alpha_{J2000} = 12 \text{ h } 36 \text{ min } 27.38 \text{ s}$, $\delta_{J2000} = +62^\circ 17' 55.56''$); it is detected in the R-band and filterless (clear) STIS image, but is undetected in the J-band image. The bright galaxy northwest of STIS123627+621755 shows an emission line at 9,660 Å, most probably identified with either [O II] $\lambda = 3,727$ Å at $z = 1.592$ or H α at $z = 0.472$.

with the spectral energy distribution reported by ref. 1, which predicts a near-infrared magnitude approximately twice as bright. For the 9,700 Å flux density reported from the slitless spectrum, our 1.2-μm non-detection implies an extremely blue continuum slope, $\alpha_{UV} > 4.7$ (95% confidence limit) where $f_\nu \propto \nu^{\alpha_{UV}}$. This is significantly steeper than the ultraviolet spectra of known star-forming galaxies⁹: longward of Ly α , the average spectral slope is $\alpha_{UV} = -0.3$ and few sources are bluer than $\alpha_{UV} = 1.5$. The slope implied by our observations is also bluer than the Rayleigh-Jeans tail of a black body spectrum ($f_\nu \propto \nu^2$), requiring either photometric errors, variability, or a non-thermal origin to the flux.

Convolving the star-forming galaxy model (Fig. 2) with an R-band filter transmission and detector response, we predict that a $z = 6.68$ galaxy with $f_\nu \approx 0.67 \mu\text{Jy}$ longward of Ly α should have $R_{AB} > 32$. The R-band detection more than 100 times brighter than the prediction implies that STIS123627+621755 cannot be at $z = 6.68$. Furthermore, much of the STIS filterless imaging flux must derive from wavelengths shortward of 9,300 Å, implying that the spectrophotometric decrement at 9,300 Å must be much less extreme than suggested. This is consistent with our 1.2-μm non-detection and implies that the 1-μm spectral slope is not as extreme as discussed above.

What is the likely redshift for STIS123627+621755? The marginally resolved STIS morphology, reported emission line, red $AB(\text{clear}) - R$ colour, and blue ground-based colour argue against a Galactic origin for the source. In particular, for $R - J < 1.5$, a stellar interpretation conservatively requires a classification earlier than K5 (see Table 3.10 of ref. 10), implying a distance more than 100 kpc for luminosity class V. A faint Galactic white dwarf identification is not ruled out by the ground-based colour. If STIS123627+621755 is comparable to the lowest-luminosity white dwarf known¹¹, ESO 439-26 with $M_V \approx 17.5$, the implied distance

is about 1 kpc. Most white dwarfs are 1.5–8 magnitudes more luminous¹².

The imaging results are consistent with a flat spectrum source of approximately 27th magnitude. If we assume the isolated emission line detected at 9,337 Å is real, then $[O\text{ II}] \lambda = 3,727 \text{ Å}$ is the more likely identification and STIS123627+621755 is similar in luminosity ($M_B \approx -17$) to the Small Magellanic Cloud at $z = 1.51$ (see Fig. 2). If so, the inferred restframe $[O\text{ II}] \lambda = 3,727 \text{ Å}$ equivalent width is $W_{[O\text{ II}]} \approx 4 \text{ Å}$, typical of local samples of star-forming galaxies¹⁵. The flux decrement at 9,300 Å is reproduced by this model, albeit at a lower amplitude, and is associated with hydrogen Balmer absorption rather than hydrogen Lyman absorption. The blue $R - J$ colour and failure of our extensive spectroscopy to detect any features both support this interpretation. However, the $AB(\text{clear}) - R$ colour is about 1 magnitude too red and suggests either variability or a spectral break in the range $\lambda = 3,000\text{--}6,000 \text{ Å}$.

These results and other recent serendipitously identified sources^{13,14} show that low-luminosity, $[O\text{ II}]$ -emitting galaxies at $z \approx 1.5$ can be convincing doppelgängers of high-redshift galaxies. At $z \approx 3$ we know that the faint end of the galaxy luminosity function is steep⁵. Low-luminosity, $[O\text{ II}]$ -emitting galaxies are likely to be a significant contaminant to high-redshift Ly α searches, suggesting that such surveys should incorporate deep imaging bluewards of the emission line to distinguish the populations. □

Received 5 July; accepted 3 October 2000.

1. Chen, H.-W., Lanzetta, K. M. & Pascarelle, S. Spectroscopic identification of a galaxy at a probable redshift of $z = 6.68$. *Nature* **398**, 586–588 (1999).
2. Williams, R. E. *et al.* The Hubble Deep Field: observations, data reduction, and galaxy photometry. *Astron. J.* **112**, 1335–1389 (1996).
3. Barger, A. *et al.* Constraints on the early formation of field elliptical galaxies. *Astron. J.* **117**, 102–110 (1999).
4. Steidel, C. S., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. Spectroscopic

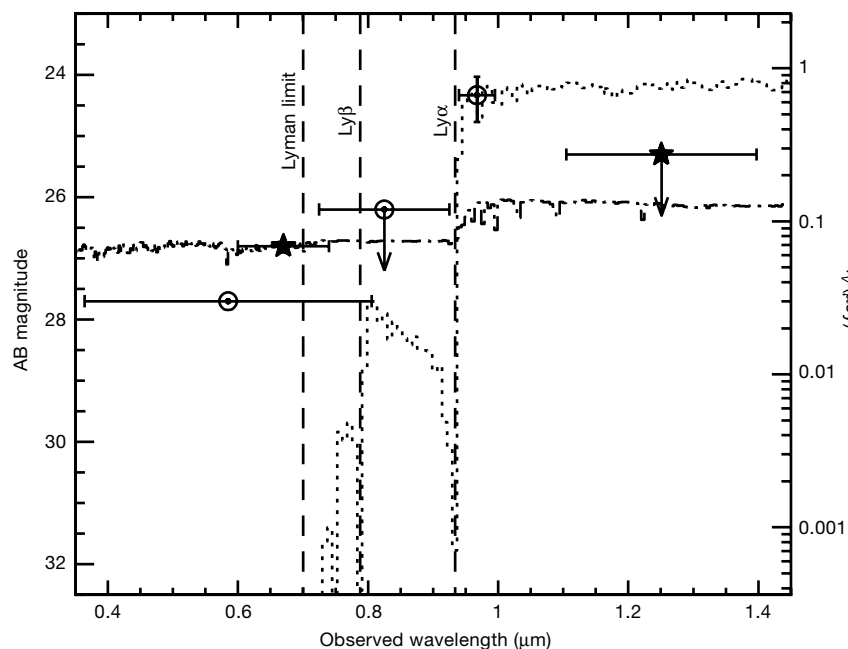


Figure 2 The photometric detection of STIS123627+621755 at 6,700 Å is not consistent with $z > 6$. Solid stars show our R- and J-band Keck photometry, where horizontal extent illustrates the width of the interference filters. We note that the R-band filter has a sharp red cutoff, unlike conventional R-band filters. Open circles are STIS imaging and spectrophotometric results from ref. 1. Arrows are 3σ upper limits. The dotted and dot-dashed lines refer to 50-Myr-old star-forming galaxy models at redshifts of $z = 6.68$ and $z = 1.51$, respectively. The model¹⁶ is a 50-Myr-old instantaneous starburst with solar metallicity, a Salpeter initial mass function from 1 to 100 solar masses, and nebular lines, attenuated by a model of the intervening intergalactic medium¹⁷. Vertical lines indicate the

strong edges due to the Ly α forest, the Ly β forest, and the Lyman limit (rest-frame 912 Å) at $z = 6.68$. The Keck photometry does not support the suggested $z = 6.68$ redshift and spectral energy distribution. In particular, the detection at 6,700 Å requires that STIS123627+621755 is not at $z > 6$. The ground-based photometric data is more consistent with the emission line reported at 9,337 Å being associated with $[O\text{ II}] \lambda = 3,727 \text{ Å}$ from a star-forming galaxy at $z = 1.51$. In this case, the luminosity of STIS123627+621755 is comparable to the Small Magellanic Cloud at a lookback time of approximately 10 Gyr.

- confirmation of a population of normal star-forming galaxies at redshifts $z < 3$. *Astrophys. J.* **462**, L17–L21 (1996).
5. Steidel, C. S., Adelberger, K. L., Gialavisco, M., Dickinson, M. & Pettini, M. Lyman-break galaxies at $z \approx 4$ and the evolution of the ultraviolet luminosity density at high redshift. *Astrophys. J.* **519**, 1–17 (1999).
 6. Stern, D. *et al.* Discovery of a color-selected quasar at $z = 5.50$. *Astrophys. J.* **533**, L75–L78 (2000).
 7. Matthews, K. & Soifer, B. T. In *Infrared Astronomy with Arrays: The Next Generation* (ed. McLean, I.) 239–246 (Kluwer, Dordrecht, 1994).
 8. Oke, J. B. *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107**, 375–385 (1995).
 9. Adelberger, K. L. & Steidel, C. C. Constraints on dusty star formation at high redshift from ultraviolet, far-infrared, and radio surveys. *Astrophys. J.* (in the press); preprint astro-ph/0001126 at (xxx.lanl.gov) (2000).
 10. Binney, J. & Merrifield, M. *Galactic Astronomy* (Princeton Univ. Press, Princeton, 1998).
 11. Ruiz, M., Bergeron, P., Leggett, S. & Anguita, C. The extremely low luminosity white dwarf ESO 439–26. *Astrophys. J.* **455**, L159–L162 (1995).
 12. Harris, H. *et al.* A very low luminosity, very cool, DC white dwarf. *Astrophys. J.* **524**, 1000–1007 (1999).
 13. Stern, D., Bunker, A. J., Spinrad, H. & Dey, A. One-line redshifts and searches for high-redshift Ly α emission. *Astrophys. J.* **537**, 73–79 (2000b).
 14. Stockton, A. & Ridgway, S. E. Deep spectroscopy in the field of 3C 212. *Astron. J.* **115**, 1340–1347 (1998).
 15. Hogg, D. W., Cohen, J. G., Blandford, R. & Pahre, M. A. The [O II] luminosity density of the universe. *Astrophys. J.* **504**, 622–628 (1998).
 16. Leitherer, C. *et al.* Starburst99: synthesis models for galaxies with active star formation. *Astrophys. J. Suppl.* **123**, 3–40 (1999).
 17. Madau, P. Radiative transfer in a clumpy universe: the colors of high-redshift galaxies. *Astrophys. J.* **441**, 18–27 (1995).

Acknowledgements

We acknowledge H. W. Chen and K. Lanzetta, who have been generous with information and supportive of our follow-up efforts on this intriguing source. We thank J. Gardner and J. Bloom for comments on the STIS data. We thank the staff of Keck Observatory for their help in obtaining the data. The W.M. Keck Observatory is operated as a scientific partnership among the University of California, the California Institute of Technology, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. We especially thank G. Punawai and X. Jerome for their assistance during the observing runs. The work of D.S. and P.E. were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. The work of W.v.B. and W.d.V. at the Institute of Geophysics and Planetary Physics, Lawrence Livermore National Laboratory was performed under the auspices of the US Department of Energy by University of California Lawrence Livermore National Laboratory. This work has been supported by a grant from the NSF (H.S.).

Correspondence and requests for materials should be addressed to D.S. (e-mail: stern@zwoelfkinder.jpl.nasa.gov).

Unusual spectral energy distribution of a galaxy previously reported to be at redshift 6.68

Hsiao-Wen Chen*, Kenneth M. Lanzetta†, Sebastian Pascarelle† & Noriaki Yahata†

* Observatories of the Carnegie Institution of Washington, 813 Santa Barbara, Pasadena, California 91101, USA

† Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, New York 11794-3800, USA

Observations of distant galaxies are important both for understanding how galaxies form and for probing the physical conditions of the Universe at early times. It is, however, very difficult to identify galaxies at redshifts $z > 5$, because they are so faint and have few spectral characteristics. We previously reported¹ the probable identification of a galaxy at $z = 6.68$, based on one line and an apparent break in the spectrum just shortwards of that, which we interpreted as Lyman α emission and the Lyman α break, where photons with shorter wavelengths are absorbed by the intervening neutral hydrogen gas. Here we present optical photometry that shows moderate detections of light in the B- and

V-band images, which are inconsistent with the expected absence of flux shortwards of the Lyman α break for a galaxy at $z > 5$, and inconsistent with the previous flux measurement. Moreover, the spectral energy distribution for this object cannot readily be fitted by any known galaxy spectral template at any redshift, so the redshift is undetermined.

The galaxy (designated galaxy A, ref. 1) was previously identified in deep observations obtained by the Hubble Space Telescope using the Space Telescope Imaging Spectrograph (STIS) in December 1997. It has a STIS clear magnitude of $AB(\text{clear}) = 27.67 \pm 0.09$. The spectrum of galaxy A was extracted using a new spectrum extraction technique that employed optimal weights to spatially deblend and spectrally deconvolve the slitless spectra. The spectrum was characterized by a discontinuity at wavelength $\lambda \approx 9,300 \text{ \AA}$, which we interpreted as the Lyman- α decrement (produced by intervening hydrogen absorption), and by an emission line at wavelength $\lambda \approx 9,334$, which we interpreted as Lyman- α . But because of the low signal-to-noise ratios of the identified spectral features, the redshift identification of galaxy A was not indisputable.

It is extremely difficult to obtain a confirming spectrum even with a 10-m class telescope on the ground because, at near-infrared wavelengths, background sky light is the dominant source of noise and confusing sky emission lines make identifications of spectral line features ambiguous. To strengthen the identifications of the faint galaxies observed in the STIS field, we acquired deep optical B, V and R images of the STIS field with the WIYN (Wisconsin–Indiana–Yale–NOAO (National Optical Astronomy Observatories)) telescope at the Kitt Peak National Observatory using the Mini Mosaic camera. The primary objective was to obtain sensitive photometry over a wide passband in order to provide strong constraints in the spectral energy distributions of these galaxies. In particular, a complete absence of flux at optical wavelengths shortwards of the redshifted Lyman- α emission line owing to intervening neutral hydrogen absorption would unambiguously confirm the previous redshift identification of galaxy A. The observations were carried out in a series of exposures of between 900 and 1,200 s each in January 2000. The total exposure times of the B, V and R images were 2.3, 2.0 and 3.5 h, which reached 5σ point-source-limiting AB magnitudes of about 26.3, 26.0 and 26.7, respectively. The individual exposures were reduced using standard pipeline techniques and were registered to a common origin and filtered for cosmic rays.

To measure the energy fluxes of galaxies in the STIS field, we adopted a quasi-optimal photometry technique that fits model spatial profiles of detected objects to the individual ground-based images². The model spatial profiles were obtained by convolving smooth models of the intrinsic profiles of detected objects (determined on the basis of the direct image previously obtained with STIS using a non-negative least-squares image reconstruction method³) with appropriate point spread functions of individual images (approximated using a gaussian profile). The new technique provides higher signal-to-noise ratio and more accurate flux uncertainty measurements than standard techniques, because it properly accounts for the variation of the point spread functions of individual frames (the full-width at half-maximum of which was

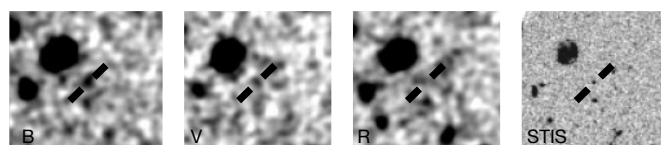


Figure 1 Portions of the ground-based B, V and R images, smoothed by the size of the point spread functions, and the corresponding STIS image centred at galaxy A. Each panel is 14 arcsec on a side. North is to the left and East is down. The AB magnitudes of galaxy A are $AB(B) = 26.7 \pm 0.2$, $AB(V) = 26.8 \pm 0.4$, $AB(R) = 27.3 \pm 0.4$, and $AB(\text{clear}) = 27.67 \pm 0.09$.

measured to vary between 0.5 and 1.0 arcsec) and because it accounts for uncertainty correlations between nearby, overlapping neighbours.

We present the new ground-based B, V and R images of galaxy A in Fig. 1 together with the direct image obtained in the STIS clear filter. Contrary to our expectations on the basis of the previous analysis, galaxy A exhibits moderate detections of flux in all three ground-based images. Specifically, we estimate that galaxy A has B, V and R magnitudes of $AB(B) = 26.7 \pm 0.2$, $AB(V) = 26.8 \pm 0.4$ and $AB(R) = 27.3 \pm 0.4$, respectively. For a galaxy at redshift $z = 6.68$, both the B and V filters cover a wavelength range below the redshifted Lyman- α emission line. Therefore, the presence of optical flux in the B and V images indicates that galaxy A is very unlikely to be at redshift $z = 6.68$. Nevertheless, it is puzzling that the ground-based photometry do not seem to agree with the previous STIS photometry. We find that for a STIS clear magnitude of $AB(\text{clear}) = 27.67 \pm 0.09$, galaxy A is at least a factor of two too bright in the B and V filters when compared with known galaxy spectral energy distributions of any given redshift.

Figure 2 shows the spectral energy distribution of galaxy A established from the broad-band photometry in the B, V, R (open circles) and STIS clear filters (open square) in comparison to the extracted one-dimensional STIS spectrum of galaxy A cast into 350 Å bins (open triangles). The spectral coverage of each filter is indicated by horizontal bars. It further shows that galaxy A has an extremely blue colour, $AB(B) - AB(R) = -0.6 \pm 0.4$. The only way that the new photometry could be in line with the previous measurement would be if a dominant portion of the flux observed in the STIS clear filter occurred within the spectral range covered by the B, V and R filters (that is, between 3,500 and 9,000 Å) and a negligible amount of flux at wavelengths below or beyond this range. The closest resemblance would be the spectral energy distribution of an F-type star, but the apparent magnitude of this object would imply a distance of about 450 kpc from the Sun, many times larger than the size of the Milky Way. We therefore conclude that this cannot be the case.

Because the STIS direct image (covering a wavelength range that spans from 2,000 Å to 10,000 Å) is a factor of approximately 10 times more sensitive than the ground-based images and because STIS photometric calibration is accurate to within 5% (according to the STIS Instrument Handbook), it ought to provide a strong, reliable constraint in the integrated optical flux of galaxy A. To resolve the discrepancy between the ground-based and STIS photometry for galaxy A, we first examined the photometric calibration of

the ground-based images. First, we measured photometry in the B, V and R images for all the 1,067 objects in the Hubble Deep Field (HDF) region presented by ref. 4. Next, we converted the ground-based B, V and R photometry to the space-based F450W and F606W photometry using the transformation presented in ref. 5. Finally, we compared our photometry of these objects with the ones presented in ref. 4. We found that these measurements agreed with each other. We therefore concluded that the photometric calibration of the ground-based images was consistent with the space-based photometry and that the discrepant photometry was owing to some other factor. Because the two sets of observations were taken two years apart, the discrepant photometry could suggest that galaxy A is variable. In particular, the observed blue $V - R$ colour of galaxy A agrees with that of the supernova 1995K identified at $z = 0.479$ (ref. 6). But we cannot provide further support to this proposal.

Regardless of the peculiarity of the spectral energy distribution of galaxy A, the conflicting results from two separate analyses also suggest that the previous analysis of the STIS observations might be in error. We therefore carefully examined the previous analysis, particularly the spectral extraction technique.

The primary arguments that were presented by ref. 1 to support the confidence of the redshift identification were: (1) consistent flux measurements from the direct image ($AB(\text{clear}) = 27.67 \pm 0.09$) and the dispersed image ($AB(\text{clear}) = 27.72 \pm 0.29$); (2) a robust detection of the emission line at $\lambda \approx 9,334$ Å; and (3) a statistically significant detection of the flux discontinuity at $\lambda \approx 9,330$ Å that coincided with the emission line. We confirmed the detection of the emission line and the flux discontinuity using traditional spectral extraction techniques, but we were unable to confirm the amplitude of the flux discontinuity. The amplitude of the flux discontinuity could be overestimated in the previous analysis due to small systematic errors in the sky determination of the dispersed image, but we have not found such evidence.

The results of our analysis show that galaxy A exhibits unusual spectral properties that cannot be explained by either galaxies or stars. We cannot resolve the incompatible photometry of the new ground-based images and the previous STIS images. Therefore, we conclude that the redshift identification of galaxy A remains undetermined. □

Received 26 June; accepted 5 October 2000.

1. Chen, H.-W., Lanzetta, K. M. & Pascarelle, S. A spectroscopically identified galaxy of probable redshift $z = 6.68$. *Nature* **398**, 586–588 (1999).
2. Yahata, N. *et al.* Photometry and photometric redshifts of faint galaxies in the Hubble deep field south NICMOS field. *Astrophys. J.* **538**, 493–504 (2000).

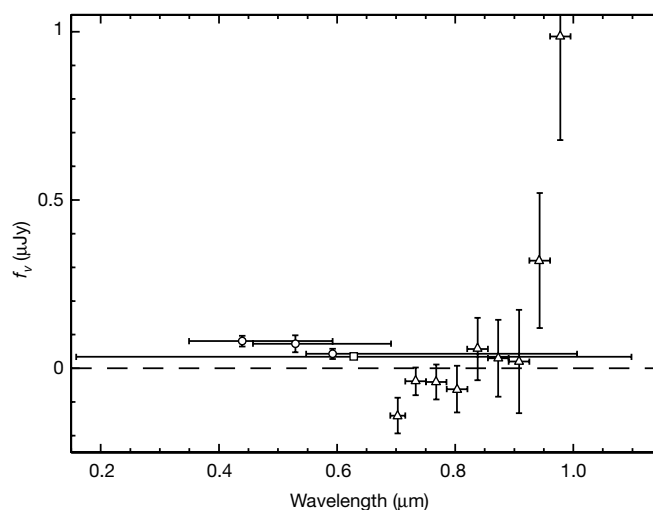


Figure 2 The observed spectral energy distribution of galaxy A. The open circles represent the broad-band photometry in the B, V and R images from the WIYN telescope. The open square represents the photometric measurement in the STIS clear filter. The

open triangles represent the extracted one-dimensional STIS spectrum of galaxy A cast into 350 Å bins. The spectral coverage of each filter is indicated by horizontal bars.

3. Puetter, R. C. & Yahil, A. in *The Astronomical Data Analysis Software and Systems (ADASS) VIII Conference* (eds Mehringer, D. M., Plante, R. L. & Roberts, D. A.) Vol. 172, 307–316 (Astronomical Society Publications, ASP, San Francisco, 1999).
4. Fernández-Soto, A., Lanzetta, K. M. & Yahil, A. A new catalog of photometric redshifts in the Hubble deep field. *Astrophys. J.* **513**, 34–50 (1999).
5. Holtzman, J. A. *et al.* The photometric performance and calibration of WFPC2. *Publ. Astron. Soc. Pacif.* **107**, 1065–1093 (1995).
6. Schmidt, B. P. *et al.* The high-*z* supernova search: measuring cosmic deceleration and global curvature of the universe using type Ia supernovae. *Astrophys. J.* **507**, 46–63 (1998).

Correspondence should be addressed to H.-W.C. (e-mail: hchen@ociw.edu).

First-order transition in confined water between high-density liquid and low-density amorphous phases

Kenichiro Koga^{*}, Hideki Tanaka[†] & X. C. Zeng[‡]

^{*} Department of Chemistry, Fukuoka University of Education, Fukuoka 811-4192, Japan

[†] Department of Chemistry, Okayama University, 3-1-1, Tsushima, Okayama 700-8530, Japan

[‡] Department of Chemistry and Center for Materials & Analysis, University of Nebraska, Lincoln, Nebraska 68588, USA

Supercooled water and amorphous ice have a rich metastable phase behaviour. In addition to transitions between high- and low-density amorphous solids^{1,2}, and between high- and low-density liquids^{3–8}, a fragile-to-strong liquid transition has recently been proposed^{9,10}, and supported by evidence from the behaviour of deeply supercooled bilayer water confined in hydrophilic slit pores¹¹. Here we report evidence from molecular dynamics simulations for another type of first-order phase transition—a liquid-to-bilayer amorphous transition—above the freezing temperature of bulk water at atmospheric pressure. This transition occurs only when water is confined in a hydrophobic slit pore^{12–14} with a width of less than one nanometre. On cooling, the confined water, which has an imperfect random hydrogen-bonded network, transforms into a bilayer amorphous phase with a perfect network (owing to the formation of various hydrogen-bonded polygons) but no long-range order. The transition shares some characteristics with those observed in tetrahedrally coordinated substances such as liquid silicon^{15,16}, liquid carbon¹⁷ and liquid phosphorus¹⁸.

A popular model of water—the TIP4P model—was employed¹⁹

in the molecular dynamics (MD) computer simulation. The melting temperature of the three-dimensional (3D) TIP4P normal ice (*I_h*) at ambient pressure (0.1 MPa) was recently estimated to be about 237 ± 7 K (ref. 20). We prepared a thin film of TIP4P water by confining the water molecules between two parallel walls fixed in a space. The wall–wall separation was taken to be just enough to accommodate two molecular layers. The water–wall interaction was modelled by a 9-3 Lennard–Jones potential (hydrophobic wall)¹³. The MD simulations were performed under the condition of fixed temperature and fixed lateral pressure (pressure tensor components parallel to the walls), using a modified Nosé–Andersen method. In the first series of MD simulations, a system of 960 molecules (system A) was cooled from 400 K to 210 K in steps and then heated back to 400 K, at a lateral pressure of 0.1 MPa. Three independent MD simulations were carried out with this cooling-and-heating process. In the second series, a system of 240 molecules (system B) was used to examine the phase behaviour at several fixed pressures: 0.1 MPa, 100 MPa, 300 MPa and 1 GPa. At each pressure the temperature was, again, lowered in steps. The MD simulation time ranged from 200 ps to 20 ns, depending on the temperature and pressure of the system. For the sake of structure analysis, configurations of system A at 275, 270 and 210 K were mapped respectively onto the inherent structures²¹ by means of the steepest descent method at a fixed volume.

Figure 1 shows the temperature dependence of the potential energies at a fixed pressure of 0.1 MPa. As the system is cooled, the potential energy first gradually decreases and then suddenly drops by about 5.6 kJ mol^{-1} to a lower value at 270 K. This large energy drop within a small temperature range (5 K) is comparable to that (7.0 kJ mol^{-1} , ref. 20) at the freezing transition of TIP4P bulk water and is much larger than that (1.2 kJ mol^{-1} at 200 MPa, ref. 6) at the high-density liquid (HDL) to low-density liquid (LDL) transition in supercooled TIP4P bulk water. In the reversed heating process, the potential energy jumps to a higher value at around 300 K, following the trace of the cooling process. The sharp changes in energy and the large hysteresis appearing in the cooling and heating processes suggest a strong first-order phase transition in the confined water. The temperature dependence of the density of the system also shows that the transition is first-order, as the density initially increases gradually with lowering the temperature but then decreases steeply at the transition temperature (by about 1% at 0.1 MPa and 4% at 300 MPa). Moreover, the diffusion constants calculated before and after the transition (at 275 K and 270 K respectively) differ by four orders of magnitude. The mechanism is also changed at this temperature from a free diffusion to a jump diffusion, which is highly cooperative but involves very few molecules. This order-of-magnitude change in diffusivity is comparable to that of the liquid-to-crystalline ice transition in the 3D TIP4P system (Table 1), as well

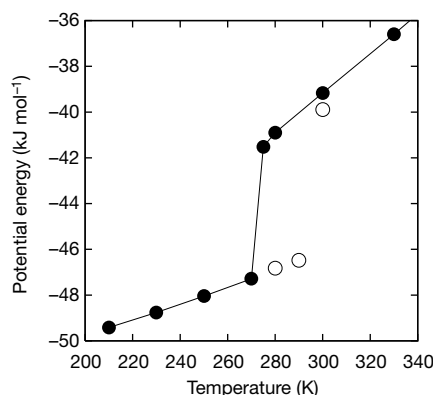


Figure 1 Temperature dependence of the potential energy. The potential energy consists of the water–water intermolecular interactions and the water–wall interactions. Filled and unfilled circles indicate the cooling and heating processes, respectively.

as to that of real water at the liquid-to-crystalline ice transition^{22–24}. The drastic change in diffusivity within a small temperature range (5 K) suggests that the confined liquid water transforms into a very viscous ‘solid-like’ amorphous phase around 270 K, 33 K above the melting temperature of the TIP4P ice at 0.1 MPa. In contrast, only about one order of magnitude is the diffusivity change in TIP4P bulk water within a temperature range (213–233 K at 0.1 MPa and 173–183 K at 200 MPa) of a possible HDL–LDL transition (Table 2).

Shown in Fig. 2a and b are inherent structures of the liquid and the new bilayer amorphous phase, respectively. In the amorphous phase, the two molecular layers are nearly flat; the spatial arrangements of oxygen atoms in both layers are being superimposed on each other (that is, they have a symmetry of reflection). These two features do not show up in the liquid phase. More crucial is the difference in the hydrogen-bond network structure. In the new bilayer amorphous phase, every molecule is hydrogen-bonded with four nearest-neighbour molecules (except for a few defects) so that three neighbours are in the same layer and the fourth one is in the opposite layer. Consequently, the hydrogen-bond network of the amorphous phase is almost perfect. Indeed, the substantial decrease in potential energy as well as the loss of fluidity upon the phase transition can be attributed mainly to the formation of the nearly perfect bilayer hydrogen-bond network. In passing, we point out that the observed liquid-to-bilayer amorphous transition occurs only in the two-layer water system. We have also examined a three-layer water system in a thicker nanopore. No liquid-to-bilayer amorphous transition was seen at 0.1 MPa.

Although the bilayer amorphous phase has a local network structure very similar to that of the bilayer crystalline ice¹³, it lacks long-range order in the configuration of oxygen atoms, as shown in the radial distribution function (Fig. 2d). There is no translational periodicity in the amorphous phase as there is in the liquid phase. Also, there is a significant spread in hydrogen-bond angles in the two-dimensional (2D) network (Fig. 2e). In contrast, the bilayer crystalline ice does have a periodic unit cell as well as three definite bond angles (Fig. 2c and e). The most notable structural difference between the bilayer amorphous and the bilayer crystalline ice is the variation in the shape of the polygonal rings of hydrogen bonds in each layer. Figure 2b shows that there are not only hexagonal rings but also pentagonal and heptagonal rings, and even some octagonal rings. The distribution of the *n*-membered rings, which was evaluated on the basis of three independent sets of 100 inherent structures, shows that it is 77% hexagonal (*n* = 6), 11% pentagonal (*n* = 5), 11% heptagonal (*n* = 7) and 1% the other rings (*n* = 4, 8, 9, 10). The appearance of a large number of pentagonal and heptagonal rings indicates that these rings are not some types of

imperfection but characteristic constituents of the hydrogen-bond network. Also found is a variety of shapes within each group of *n*-membered rings, compared to the bilayer crystalline ice in which all the hexagonal rings are congruent. These differences in the network structure between the quasi-2D bilayer amorphous and the bilayer crystalline ice are similar to the fundamental differences between the 2D continuous-random network model of a glass and the regular honeycomb network model of a crystal²⁵.

Existence of the liquid–liquid and/or amorphous–amorphous transition in supercooled bulk water is still controversial^{13–5,7,24,26–28}. Here, our simulation directly demonstrates that water, when confined in a hydrophobic nanopore, can exhibit a sharp transition to a bilayer amorphous phase (bilayer polyamorphism) at a temperature above the melting temperature of the 3D TIP4P ice. In contrast, the HDL–LDL liquid–liquid transition in bulk water is, if it exists, below 210 K (refs 5 and 7). Moreover, the bilayer polyamorphic transition shows a marked change in potential energy, density and diffusion constant within a very small temperature range (5 K). These characteristics of first-order transition are similar to those found in covalent tetrahedral substances²⁹ including liquid silicon¹⁶, carbon¹⁷, and phosphorus¹⁸.

The relation among the quasi-2D, high-density liquid (bilayer HDL), low-density amorphous (bilayer LDA), and crystalline (bilayer ice) phases can be compared with that among three corresponding 3D HDL, LDL, and crystalline ice (*I_h*) phases (Table 2): (1) the potential energy differs more strongly between the higher-energy liquid phase and the lower-energy amorphous phase than between the lower-energy amorphous phase and the crystalline phase in inherent structure, a behaviour found in both 2D and 3D systems. (2) At the transition, the population of molecules with coordination number other than 4 nearly vanishes in the quasi-2D system, and also decreases substantially in the 3D system.

Table 2 Analogies between the quasi-2D and 3D systems

Phase	Potential energy* (kJ mol ^{−1})		Four-hydrogen-bond species† (%)		Six-membered rings‡ (%)	
	Q2D	3D	Q2D	3D	Q2D	3D
HDL	−47.6	−53.5	73.1	80.2	35.4	46.0
LDL§	−51.9	−55.0	99.7	92.2	76.7	62.2
Crystalline	−53.0	−55.9	100.0	100.0	100.0	100.0

Q2D, quasi-two-dimensional; HDL, high-density liquid; LDL, low-density liquid; LDA, low-density amorphous.

* Average potential energy of inherent structures.

† Average fraction of molecules having four hydrogen-bonding neighbours.

‡ Average number of six-membered rings normalized by the number in a corresponding crystal.

§ Bilayer LDA in the case of the quasi-2D system.

Table 1 Diffusion constants for the TIP4P water in the quasi-2D and 3D systems (in cm² s^{−1})

Quasi-2D system				
Hydrophobic pore*		Hydrophilic pore†	Hydrophobic pore, larger width‡	
275K: 1.3×10^{-5}				
270K: 3.0×10^{-9}				
250K: 7.9×10^{-11}		250K: 5.0×10^{-7}	250K: 6.2×10^{-6}	
3D system				
Supercooled (0.1 MPa)	Supercooled (200 MPa)	Ice	Supercooled (exp.)§	Ice (exp.)
233K: 2×10^{-6}	193K: 3.0×10^{-7}	240K: 3×10^{-10}	263.7K: 7.0×10^{-6}	252K: 1.0×10^{-9}
213K: 9×10^{-8}	183K: 1.4×10^{-7}	230K: 0¶	250K: 3.4×10^{-6}	242K: 7.7×10^{-10}
193K: 6×10^{-8}	173K: 9×10^{-9}			

* The system of interest.

† A reference system of water confined in a model hydrophilic slit pore.

‡ Another reference system with a slit width 1.5 times wider than that of the system of interest.

§ Ref. 22.

|| Ref. 23.

¶ Less than 10^{-11} .

(3) In both 2D and 3D systems, the number of six-membered rings in the hydrogen-bond network differs considerably among the three phases. (4) The diffusion constant changes by at least four orders of magnitude from liquid to the crystalline ice phase or to the bilayer LDA phase (Table 1).

We also examined the phase behaviour of the confined water in a wide range of pressure and temperature with system B. We found that the higher the pressure the lower the transition temperature (265 K at 100 MPa, and 260 K at 300 MPa). This behaviour is consistent with a negative change in entropy ($\Delta S < 0$) and a positive change in volume ($\Delta V > 0$) at the liquid-to-bilayer amorphous phase transition. Moreover, the same transition was observed around 500 MPa in a decompressing process from 1 GPa to 300 MPa at a fixed temperature of 250 K. However, the transition was not observed beyond a high pressure of 1 GPa and above 190 K. The expansion ($\Delta V > 0$) at the liquid-to-bilayer amorphous transition is analogous to that at the bulk water-to-ice freezing transition.

We note that our finding of the first-order liquid-to-bilayer amorphous transition is based on the TIP4P model of water. Confirmation of this finding must await experiments. Recently, Bergman and Swenson¹¹ studied bilayer water confined in hydrophilic nanopores using broad-band dielectric spectroscopy. They found that about half of the water molecules form a hydrogen bond to one of the wall surfaces and they also found evidence to suggest a fragile-to-strong liquid transition in deeply supercooled water¹⁰. We have performed a preliminary simulation of bilayer water confined in a model hydrophilic pore. The bilayer HDL to LDA transition was not seen within the timescale of the simulation. The diffusion constant calculated at 250 K is similar to that of bulk liquid water, that is, four orders of magnitude higher than that of the bilayer amorphous phase. In the hydrophobic pores, on the other hand, water molecules do not form hydrogen bonds to the wall surfaces and the bilayer HDL to LDA transition occurs. The question is why such a transition happens only in the hydrophobic nanopore. One explanation may be that when the two-layer water is confined in the

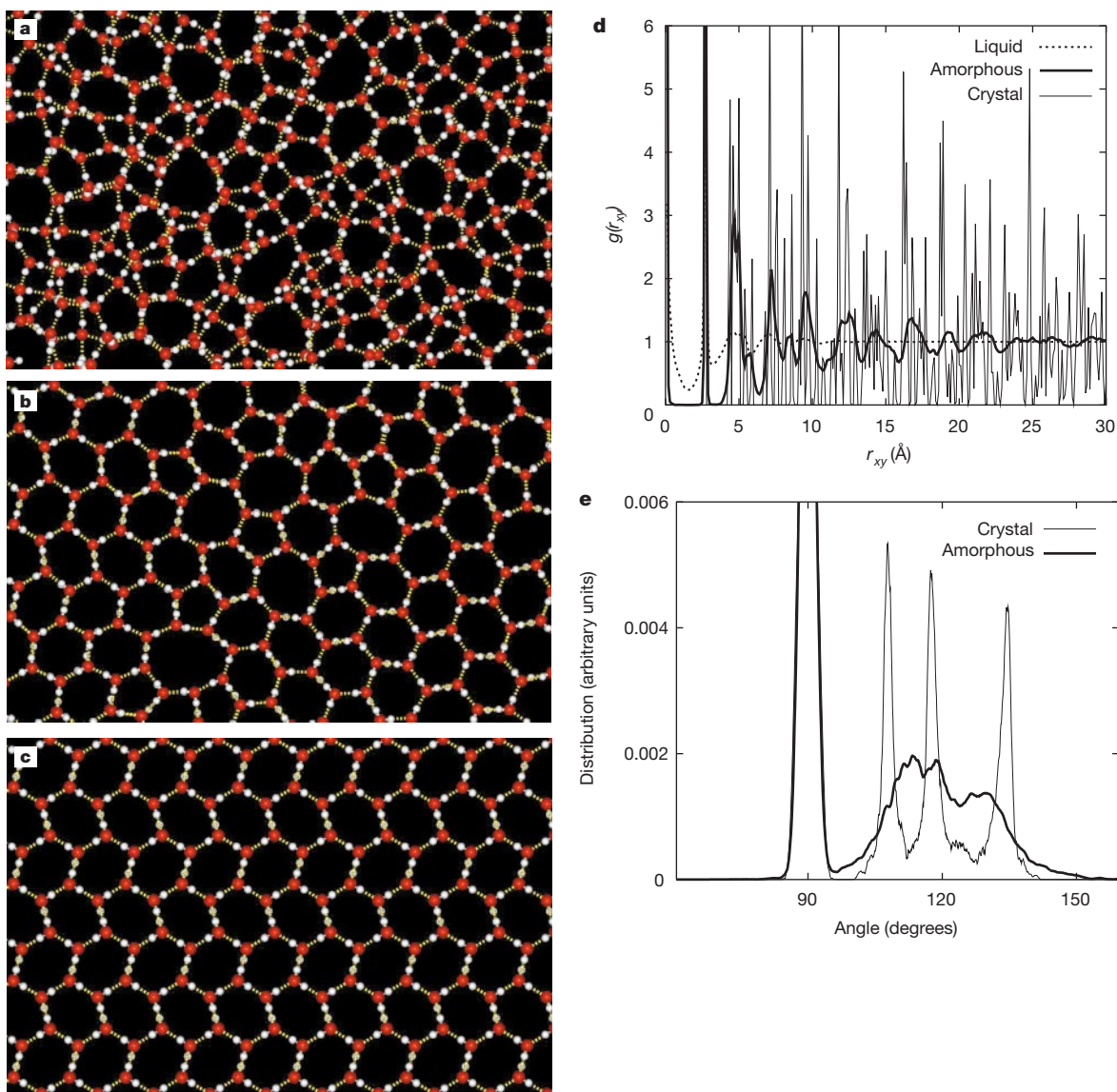


Figure 2 Structure of the confined two-layer water in liquid, amorphous, and crystalline ice phases. **a**, Liquid. **b**, Amorphous. **c**, Crystalline ice. These are top views, showing in-plane (2D) molecular configurations and hydrogen-bond networks. Red and white spheres are oxygen and hydrogen atoms, respectively. Yellow dotted lines indicate the hydrogen bonds. In **b** and **c**, the in-plane 3-connected hydrogen-bond network of one layer is

completely superimposed on that of the other layer and the fourth vertical hydrogen bonds connecting two layers are hidden. **d**, Radial distribution function of oxygen atoms. Distance r_{xy} is the distance in a plane parallel to the walls. **e**, Distribution of the hydrogen-bond angles.

hydrophobic nanopore, the geometrical confinement forces the water molecules to form a perfect (coordination number $z = 4$) 2D bilayer hydrogen-bond network even in a topologically disordered structure. However, in the 3D supercooled water the coordination number has a distribution of z -values (3% for $z = 3$ and 2% for $z = 5$) even after the transition to the lower-energy LDL phase³⁰ (before transition, 11% for $z = 3$ and 5% for $z = 5$). As a result, the LDL phase of 3D supercooled water (below 180 K) has a higher mobility than the bilayer LDA phase, even when the former is at much lower temperatures. Indeed, the diffusion constant of 3D LDL at 173 K (the deeply supercooled region) is still two orders of magnitude higher than that of bilayer LDA at 250 K. □

Received 31 March; accepted 10 October 2000.

- Mishima, O., Calvert, L. D. & Whalley, E. 'Melting' ice I at 77 K and 10 kbar: a new method of making amorphous solids. *Nature* **310**, 393–394 (1984).
- Mishima, O., Calvert, L. D. & Whalley, E. An apparently first-order transition between two amorphous phases of ice induced by pressure. *Nature* **314**, 76–78 (1985).
- Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. Phase behaviour of metastable water. *Nature* **360**, 324–328 (1992).
- Stanley, H. E. *et al.* Is there a second critical point in liquid water? *Physica A* **205**, 122–139 (1994).
- Tanaka, H. A self-consistent phase diagram for supercooled water. *Nature* **380**, 328–330 (1996).
- Tanaka, H. Phase behaviors of supercooled water: Reconciling a critical point of amorphous ices with spinodal instability. *J. Chem. Phys.* **105**, 5099–5111 (1996).
- Sciortino, F., Poole, P. H., Essmann, U. & Stanley, H. E. Line of compressibility maxima in the phase diagram of supercooled water. *Phys. Rev. E* **55**, 727–737 (1997).
- Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
- Angell, C. A. Water-II is a strong liquid. *J. Phys. Chem.* **97**, 6339–6341 (1993).
- Ito, K., Moynihan, C. T. & Angell, C. A. Thermodynamic fragility in liquids and a fragile-to-strong liquid transition in water. *Nature* **398**, 492–495 (1999).
- Bergman, R. & Swenson, J. Dynamics of supercooled water in confined geometry. *Nature* **403**, 283–286 (2000).
- Bellissent-Funel, M.-C., Chen, S. H. & Zanotti, J. M. X-ray and neutron scattering studies of the structure of water at a hydrophobic surface. *J. Chem. Phys.* **104**, 10023–10029 (1996).
- Koga, K., Zeng, X. C. & Tanaka, H. Freezing of confined water: a bilayer ice phase in hydrophobic nanopores. *Phys. Rev. Lett.* **79**, 5262–5265 (1997).
- Meyer, M. & Stanley, H. E. Liquid-liquid phase transition in confined water: A Monte Carlo study. *J. Phys. Chem. B* **103**, 9728–9730 (1999).
- Thompson, M. O. *et al.* Melting temperature and explosive crystallization of amorphous silicon during pulsed laser irradiation. *Phys. Rev. Lett.* **52**, 2360–2363 (1984).
- Angell, C. A., Borick, S. S. & Grabow, M. H. Glass transitions and first order liquid-metal-to-semiconductor transitions in 4-5-6 covalent systems. *J. Non-Cryst. Solids* **205**, 463–471 (1996).
- Glosli, J. N. & Ree, F. H. Liquid-liquid phase transformation in carbon. *Phys. Rev. Lett.* **82**, 4659–4662 (1999).
- Katayama, Y. *et al.* A first-order liquid-liquid phase transition in phosphorus. *Nature* **403**, 170–173 (2000).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Gao, G. T., Zeng, X. C. & Tanaka, H. The melting temperature of proton-disordered hexagonal ice: A computer simulation of TIP4P model of water. *J. Chem. Phys.* **112**, 8534–8538 (2000).
- Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **267**, 1935–1939 (1995).
- Gillen, K. T., Douglass, D. C. & Hoch, M. J. R. Self-diffusion in liquid water to -31°C . *J. Chem. Phys.* **57**, 5117–5119 (1972).
- Goto, K., Hondoh, T. & Higashi, A. Determination of diffusion coefficients of self-interstitials in ice with a new method of observing climb of dislocations by X-ray topography. *Jpn J. Appl. Phys.* **25**, 351–357 (1986).
- Smith, R. S. & Kay, B. D. The existence of supercooled water at 150 K. *Nature* **403**, 283–296 (1999).
- Zallen, R. in *The Physics of Amorphous Solids* 63–67 (Wiley, New York, 1998).
- Tse, J. S. *et al.* The mechanisms for pressure-induced amorphization of ice Ih. *Nature* **400**, 647–649 (1999).
- Soper, A. K. & Ricci, M. A. Structures of high-density and low-density water. *Phys. Rev. Lett.* **84**, 2881–2884 (2000).
- Scala, A., Starr, F. W., La Nave, E., Sciortino, F. & Stanley, H. E. Configurational entropy and diffusivity of supercooled water. *Nature* **406**, 166–169 (2000).
- Angell, C. A., Bressel, R. D., Hemmati, M., Sare, E. J. & Tucker, J. C. Water and its anomalies in perspective: tetrahedral liquids with and without liquid-liquid phase transitions. *Phys. Chem. Chem. Phys.* **2**, 1559–1566 (2000).
- Tanaka, H. Fluctuation of local order and connectivity of water molecules in two phases of supercooled water. *Phys. Rev. Lett.* **80**, 113–116 (1998).

Acknowledgements

K.K. and H.T. are supported by the Japan Society for the Promotion of Science (JSPS), the Japan Ministry of Education, and the Institute for Molecular Science (IMS). X.C.Z. is supported by the US NSF and Office of Naval Research (ONR), and by a JSPS fellowship.

Correspondence and requests for materials should be addressed to K.K. (e-mail: kenkoga@fukuoka-edu.ac.jp).

Changes in deep-water formation during the Younger Dryas event inferred from ^{10}Be and ^{14}C records

Raimund Muscheler*, Jürg Beer*, Gerhard Wagner* & Robert C. Finkel†

* Department of Surface Waters, EAWAG, CH-8600, Dübendorf, Switzerland

† Center for Accelerator Mass Spectrometry, and Geosciences and Environmental Technology Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

Variations in atmospheric radiocarbon (^{14}C) concentrations can be attributed either to changes in the carbon cycle¹—through the rate of radiocarbon removal from the atmosphere—or to variations in the production rate of ^{14}C due to changes in solar activity or the Earth's magnetic field². The production rates of ^{10}Be and ^{14}C vary in the same way, but whereas atmospheric radiocarbon concentrations are additionally affected by the carbon cycle, ^{10}Be concentrations reflect production rates more directly. A record of the ^{10}Be production-rate variations can therefore be used to separate the two influences—production rates and the carbon cycle—on radiocarbon concentrations. Here we present such an analysis of the large fluctuations in atmospheric ^{14}C concentrations, of unclear origin³, that occurred during the Younger Dryas cold period⁶. We use the ^{10}Be record from the GISP2 ice core⁵ to model past production rates of radionuclides, and find that the largest part of the fluctuations in atmospheric radiocarbon concentrations can be attributed to variations in production rate. The residual difference between measured ^{14}C concentrations and those modelled using the ^{10}Be record can be explained with an additional change in the carbon cycle, most probably in the amount of deep-water formation.

The causes of the variation in $\Delta^{14}\text{C}$ during the Younger Dryas climate event are the subject of current debate^{3,4,6–9}. $\Delta^{14}\text{C}$ is used as a measure of the atmospheric radiocarbon concentration, and is defined as the per mil deviation from the National Institute of Standards and Technology ^{14}C standard after correction for decay and fractionation¹⁰. During the Younger Dryas, ^{14}C exhibits a rapid increase followed by a gradual decrease. These changes have been discussed in the context of changes in oceanic deep-water formation⁴, but coupled ocean–climate models were not able to reproduce all aspects of the observed changes in $\Delta^{14}\text{C}$ (ref. 9). Using the reconstructed atmospheric ^{10}Be concentration as a proxy for solar activity in the early Holocene epoch, Goslar *et al.*³ claimed that it is possible to explain the $\Delta^{14}\text{C}$ during the Younger Dryas even without assuming changes in the deep-water formation. However, for a quantitative determination of the relative contributions of ocean circulation (or other effects related to the carbon cycle) and production-rate changes, a detailed knowledge of the ^{14}C production rate is crucial, because this rate has a large influence on the variations in atmospheric radiocarbon content¹.

Measurements of cosmogenic ^{10}Be in ice cores can be used to provide the necessary information on the ^{14}C production rate¹¹. ^{10}Be and ^{14}C are produced by the interaction of cosmic-ray particles with nitrogen and oxygen in the atmosphere. Recent production-rate calculations¹² allow us to determine the ^{14}C production rate, if the ^{10}Be production rate is known. To compare ^{10}Be and ^{14}C records it is necessary to take into account their different geochemical behaviour. ^{10}Be becomes attached to aerosols and is removed from the atmosphere after a mean residence time of 1–2 yr (ref. 13). By contrast, after its production ^{14}C becomes oxidized to $^{14}\text{CO}_2$ (ref. 2) and hence is involved in exchange processes between the atmosphere, biosphere and ocean as a constituent of the carbon

cycle. Therefore, ^{10}Be and ^{14}C are very similar regarding production effects, but completely different regarding their geochemistry. If the production rate of ^{14}C is known, it is then possible to disentangle the production and global-carbon-cycle effects governing $\Delta^{14}\text{C}$. In particular, the influence of the suggested decrease in North Atlantic Deep Water (NADW) formation¹⁴ on $\Delta^{14}\text{C}$ during the Younger Dryas could then be quantitatively assessed.

We use the ^{10}Be flux derived from ^{10}Be concentration measurements in the GISP2 ice core⁵ from Summit in Central Greenland to reconstruct the ^{14}C production rate during the period 15,000 to 9,300 yr BP (years before present, where present means AD 1950). Two questions challenge the use of this procedure. First, might not changes in climate lead to changes in the ^{10}Be flux at Summit? We find, to the contrary, that the ^{10}Be flux is independent of the climate proxy $\delta^{18}\text{O}$ (Fig. 1), indicating the absence of climate-related effects on the transport of ^{10}Be to Summit. Even though solar variability influences the ^{10}Be production rate and might also influence the climate, this possible connection is not visible as a correlation between ^{10}Be flux and $\delta^{18}\text{O}$. Second, there is some uncertainty as to the extent to which the ^{10}Be flux at the high geomagnetic latitude of Summit can be considered to be proportional to the global ^{10}Be production rate. Observations, however, strongly suggest that this is the case. Solar-induced excursions of ^{10}Be and $\Delta^{14}\text{C}$ during the Holocene are found to be quite similar⁵. In addition, the relationship between the Summit ^{36}Cl flux and the sediment-based record of geomagnetic field variations indicates that the radionuclides ^{10}Be and ^{36}Cl , which agree well at Summit¹⁵, reflect changes in mean global geomagnetic shielding¹⁶. For these reasons, and because central Greenland receives a considerable part of its precipitation from lower latitudes (a pattern that has persisted through the last ice age and into the Holocene^{17,18}), there is strong evidence that the flux of ^{10}Be is proportional to its global production rate. Based on this proportionality, it is possible to derive the global ^{14}C production-

rate history from the ^{10}Be record. We use a box-diffusion model¹⁹ to calculate $\Delta^{14}\text{C}$ using the ^{10}Be -derived ^{14}C production rates. To determine the extent to which the $\Delta^{14}\text{C}$ changes are caused by changes in the ^{14}C production rate alone, in a first step global deep-water formation was held constant during the model run from 15,000 to 9,300 yr BP. The variation in atmospheric CO_2 concentration⁹ and its effect on $\Delta^{14}\text{C}$ was considered, but on timescales shorter than 2,000 yr the influence of the slow atmospheric CO_2 variation during the last deglaciation is small.

The following discussion is divided into two parts. First, we compare $\Delta^{14}\text{C}$ and ^{10}Be in the early part of the Holocene (Preboreal), where we focus on the short-term variations in $\Delta^{14}\text{C}$. Second, we discuss the influence on $\Delta^{14}\text{C}$ of the large climate fluctuations that occurred before 11,500 yr BP. Figure 2a shows the ^{10}Be -derived $\Delta^{14}\text{C}$ on the timescales of the two Summit ice cores GISP2 and GRIP from 11,850 to 9,300 yr BP. Figure 2b shows $\Delta^{14}\text{C}$ back to 11,850 yr BP, as derived from tree-ring measurements according to the Intcal98 ^{14}C calibration curve²⁰. To focus on short-term variations, especially on the three large $\Delta^{14}\text{C}$ excursions on timescales of approximately 300 yr, we removed the long-term trend in $\Delta^{14}\text{C}$ by subtracting a 500-yr running mean. From 11,850 to 9,300 yr BP, the measured $\Delta^{14}\text{C}$ record and the $\Delta^{14}\text{C}$ record derived from the ^{10}Be flux resemble each other very closely. The agreement is especially striking because no free parameter is involved in fitting the ^{10}Be -derived $\Delta^{14}\text{C}$ to the tree-ring $\Delta^{14}\text{C}$. From this high degree of similarity we draw three conclusions.

(1) Changes in atmospheric ^{14}C concentration during the early Holocene on timescales shorter than 500 yr can be assigned almost completely to changes in the ^{14}C production rate. It is very improbable that synchronous changes in both the ^{10}Be transport to Summit and in the carbon cycle occurred with exactly the right magnitudes to produce the same amplitude in both calculated and measured $\Delta^{14}\text{C}$.

(2) The three biggest $\Delta^{14}\text{C}$ excursions, the biggest during the entire Holocene, occurred around 11,100, 10,100 and 9,400 yr BP.

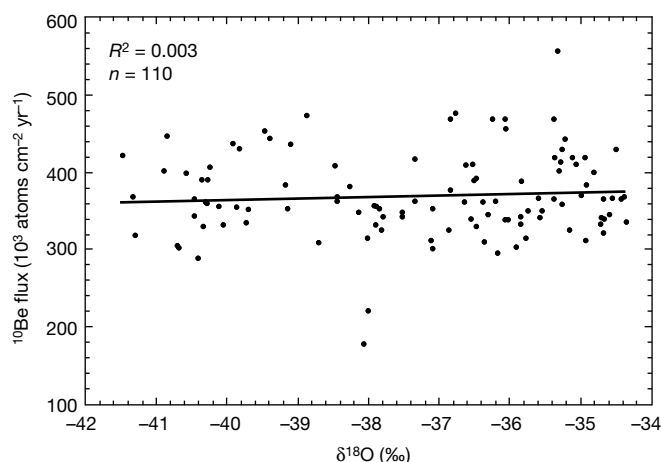


Figure 1 ^{10}Be flux versus $\delta^{18}\text{O}$ in ice at Summit. $\delta^{18}\text{O}$ denotes the relative deviation in per mil of the $^{18}\text{O}/^{16}\text{O}$ ratio from the $^{18}\text{O}/^{16}\text{O}$ value of SMOW (standard mean ocean water). Included in the figure are the ^{10}Be data from 15,000 to 9,300 yr BP. To calculate the ^{10}Be flux, the accumulation rate was averaged to match the temporal resolution of the ^{10}Be data. There are two similar approaches to reconstructing the accumulation rate. One is based on identifying the annual layers and correcting for the thinning of the layers using ice-flow models²³. The other method makes use of the relationship between accumulation rate and $\delta^{18}\text{O}$ measured in the ice core²⁷. We use the second method to calculate the ^{10}Be flux. The ^{10}Be flux at Summit is independent of $\delta^{18}\text{O}$, a climate proxy that shows large changes during the last deglaciation²⁸. This independence indicates that climate-related changes in the transport of ^{10}Be to Summit or in the deposition of ^{10}Be are lacking. Even during the rapid increase in $\delta^{18}\text{O}$ occurring at the Younger Dryas/Holocene transition, the ^{10}Be flux exhibits hardly any change, again indicating only a very small climate effect, if any.

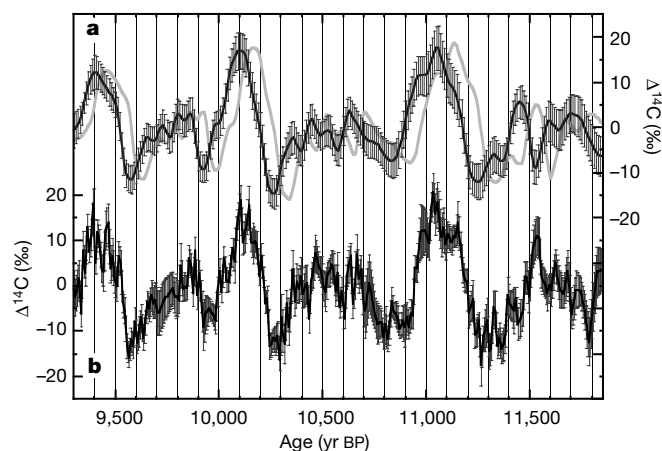


Figure 2 Comparison of modelled and measured $\Delta^{14}\text{C}$ from 11,850 to 9,300 yr BP.

a, $\Delta^{14}\text{C}$ derived from ^{10}Be measurements in the GISP2 ice core from Summit, Greenland. **b**, $\Delta^{14}\text{C}$ based on tree-ring measurements. All curves have been detrended by subtracting a 500-yr running mean. The ^{10}Be -derived $\Delta^{14}\text{C}$ is plotted on the GRIP timescale (black line) and the GISP2 timescale (grey line). To calculate $\Delta^{14}\text{C}$ based on the GRIP timescale we synchronized the $\delta^{18}\text{O}$ data sets from the GRIP²⁹ and GISP2³⁰ ice cores. A Monte Carlo simulation including 100 model runs was performed to assess the uncertainties in the production rate of ^{14}C introduced by inaccuracies in the ^{10}Be measurements of approximately 5% (ref. 5). Errors due to uncertainties in the production-rate calculations or in the reconstructed accumulation rate were not included. Tests with different detrending procedures show that longer-term differences between the detrended $\Delta^{14}\text{C}$ curves are less than 15%.

They were due to a varying geomagnetic field or to changes in solar activity. To our knowledge there are no palaeomagnetic data showing magnetic field variations that could cause these events. On the other hand, these production-rate maxima are very similar to those associated with solar-activity minima such as the Maunder minimum²¹, the only difference being that their duration is roughly twice as long. If the Preboreal maxima were caused by reduced solar activity, one might expect effects on climate. Although no obvious climate effects are visible in the $\delta^{18}\text{O}$ records from GRIP and GISP2,

there are two well-documented cold periods at 11,250 yr BP (ref. 7) and 10,300 yr BP (ref. 22) that are possibly related to the ^{10}Be concentration increase.

(3) The three production-rate maxima allow us to date the GRIP and GISP2 ice cores very precisely. Cross-correlation calculations show that the timescale of the GRIP ice core and the tree-ring chronology agree well, within the time resolution of the ^{10}Be measurements. The end of the Younger Dryas in the GRIP ice core at 11,550 yr BP seems to be very well dated, agreeing well with other chronologies³. Therefore the GISP2 timescale has to be shifted approximately 60 yr towards younger ages, which is still within the errors of this timescale²³.

To extend the ^{10}Be and ^{14}C comparison further back to the Younger Dryas and before, we use ^{14}C measurements from sediment core PL07-56PC from the Cariaco basin⁴; these yield values of $\Delta^{14}\text{C}$ during the Younger Dryas beyond the tree-ring-based calibration. The relative reflectivity (grey scale) of this core can be interpreted as a proxy for climate variability²⁴. Therefore, comparing the Cariaco basin grey scale data with the $\delta^{18}\text{O}$ values from the GRIP ice core allows a synchronization of the two records. Figure 3 shows grey scale (Fig. 3a) and $\delta^{18}\text{O}$ (Fig. 3f) together with the measured $\Delta^{14}\text{C}$ (Fig. 3b) and the ^{10}Be -based $\Delta^{14}\text{C}$ (Fig. 3d) on their respective timescales. Both measured and ^{10}Be -derived $\Delta^{14}\text{C}$ data show a similar decrease during the Younger Dryas. The $\Delta^{14}\text{C}$ decrease seems, therefore, to have been induced mainly by a changing production rate rather than by an early re-establishment of global deep-water formation during the Younger Dryas. In the light of the assumed connection between NADW formation and $\delta^{18}\text{O}$ in Greenland, this conclusion is supported by the low $\delta^{18}\text{O}$ values throughout the Younger Dryas as measured in ice cores from Summit.

At present it cannot be decided whether the changes in the ^{14}C production rate were caused by changes in solar activity or by a changing geomagnetic field. This is because, on centennial to millennial timescales, the reconstruction of the Earth's magnetic dipole field from sedimentary records is still problematic²⁵. Possible short-term variations in the geomagnetic dipole field are not well known. The increase in $\Delta^{14}\text{C}$ at the beginning of the Younger Dryas is stronger in the measured data (Fig. 3b) than in the ^{10}Be -derived data (Fig. 3d). This difference points to a change in deep-water formation and a resulting reduced ^{14}C uptake by the oceans. The grey curve in Fig. 3c shows a model run where the global deep-water formation has been reduced during the Younger Dryas, corresponding in our model to a 30% reduction in the diffusive carbon exchange in the ocean. The ^{10}Be -derived curve agrees better with the measured curve when this reduction in global deep-water formation during the Younger Dryas is included. To determine the extent to which this result depends on our adopted accumulation rate, we then used the accumulation rate based on layer counting²³ and reconstructed the atmospheric ^{10}Be concentration²⁶, which yielded a similar modelled $\Delta^{14}\text{C}$. The predicted increase in ^{10}Be -based $\Delta^{14}\text{C}$ at the beginning of the Younger Dryas was then smaller and the decrease during the Younger Dryas was even stronger compared to the measured $\Delta^{14}\text{C}$, which leads to a similar conclusion about the change in deep-water formation during the Younger Dryas.

Before 13,500 yr BP the curves based on ^{10}Be and the measured $\Delta^{14}\text{C}$ curves (Fig. 3c) do not agree. However, new $\Delta^{14}\text{C}$ data from the Cariaco basin agree more closely with the ^{10}Be data (K. Hughen, personal communication). There are several possible explanations for discrepancies between ^{10}Be and $\Delta^{14}\text{C}$. Changes in the transport of ^{10}Be to Summit would yield differences between ^{10}Be -derived and measured $\Delta^{14}\text{C}$. Discrepancies between reconstructed and actual accumulation rates would lead to errors in the calculation of the ^{10}Be flux. Changes in the carbon cycle are unlikely to produce differences between ^{10}Be -derived and measured $\Delta^{14}\text{C}$ larger than during the Younger Dryas, when the strongest climate changes took place. A different relative timing between Cariaco basin and

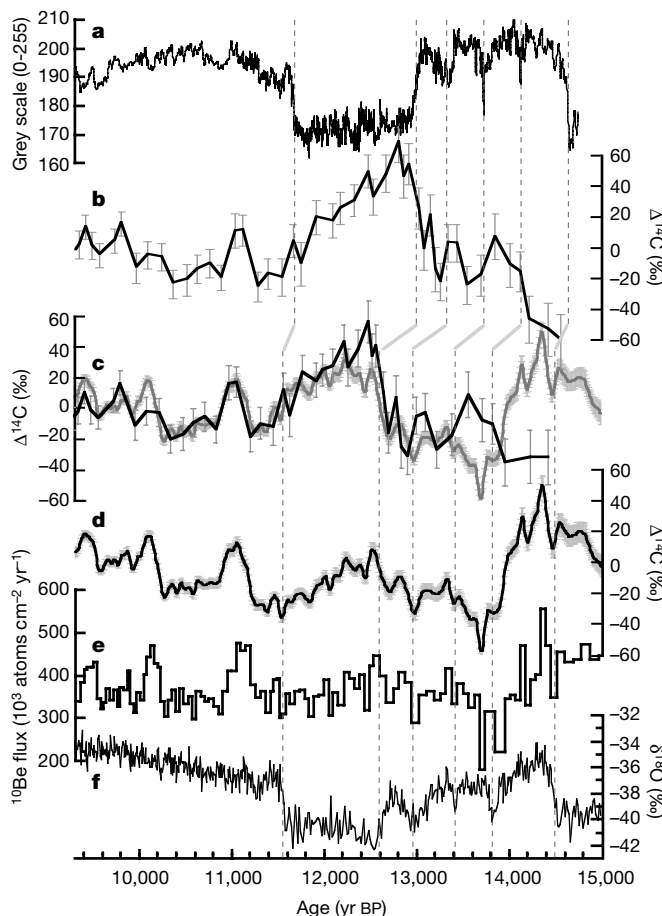


Figure 3 Comparison of modelled and measured $\Delta^{14}\text{C}$, together with ^{10}Be flux and proxy climate data from 15,000 to 9,300 yr BP. **a**, Grey scale versus age for Cariaco basin core PL07-56PC (ref. 24). The Younger Dryas cold climate event is clearly visible as a minimum in the grey scale. In the GRIP ice core the Younger Dryas is dated by annual layer counting to lie between 11,550 $\pm$ 70 and 12,700 $\pm$ 100 yr BP (ref. 28). The timescale of core PL07-56PC differs from the GRIP timescale by $\sim$ 300 yr at the beginning of the Younger Dryas and by $\sim$ 100 yr at the end. **b**, Measured $\Delta^{14}\text{C}$ values from core PL07-56PC (ref. 4) detrended by subtracting a linear fit from 15,000 to 9,300 yr BP and plotted on the same time scale as the grey scale. **c**, Measured (black line) and ^{10}Be -derived $\Delta^{14}\text{C}$ including changes in deep-water formation (grey curve). A 30% reduction in the diffusion constant throughout the Younger Dryas was assumed for the model run. This reduction in the oceanic ventilation rate produces a change in $\Delta^{14}\text{C}$ of approximately 40‰, which is comparable to model calculations with a halved rate of NADW formation⁶. As there are indications that the start of the Younger Dryas at 12,700 yr BP is well dated in the GRIP ice core³, both measured and ^{10}Be -derived $\Delta^{14}\text{C}$ are plotted on the GRIP timescale and linearly detrended. **d**, Modelled and detrended $\Delta^{14}\text{C}$ based on the ^{14}C production rate derived from the ^{10}Be flux to Summit (**e**) and assuming no changes in deep-water formation throughout the complete deglaciation. The relative timing between the climate proxies $\delta^{18}\text{O}$ measured in the GRIP ice core (**f**) and grey scale as suggested by Hughen *et al.*²⁴ is indicated by the vertical dashed lines. The correlation between measured and ^{10}Be -based $\Delta^{14}\text{C}$ from 13,500 to 10,500 yr BP increases from $R^2 = 0.47$ to $R^2 = 0.73$ ($n = 28$) if changes in deep-water formation are included.

Greenland climate would yield differences between ^{10}Be -derived and measured $\Delta^{14}\text{C}$. However, ^{14}C and ^{10}Be measurements with a better time resolution are necessary to correlate production-rate changes of these radionuclides on shorter timescales. During shorter but stable climate periods, when system changes can be excluded, we could even detect similarities in the fine structure of the two records.

^{10}Be in ice cores provides an independent tool to check the dating of ^{14}C time series older than 11,500 yr; this should allow various ^{14}C data sets to be synchronized, thus allowing a reliable extension of the tree-ring calibration curve. Additionally, high-resolution ^{10}Be data sets from other ice cores (for example, from Antarctica) could help to correct for potential transport effects on the ^{10}Be flux, and to establish a detailed production history of cosmogenic nuclides. The ^{10}Be -derived ^{14}C production rate can help in obtaining a consistent picture of the climate changes that occurred during the last deglaciation and of the role played by global deep-water formation in causing these changes. □

Received 9 March; accepted 17 October 2000.

- Siegenthaler, U., Heimann, M. & Oeschger, H. ^{14}C variations caused by changes in the global carbon cycle. *Radiocarbon* **22**, 177–191 (1980).
- Lal, D. & Peters, B. in *Handbuch für Physik* (ed. Flügge, S.) 551–612 (Springer, Berlin, 1967).
- Goslar, T., Arnold, M., Tisnerat-Laborde, N., Czernik, J. & Więckowski, K. Variations of Younger Dryas atmospheric radiocarbon explicable without ocean circulation changes. *Nature* **403**, 877–880 (2000).
- Hughen, K. *et al.* Deglacial changes in ocean circulation from an extended radiocarbon calibration. *Nature* **391**, 65–68 (1998).
- Finkel, R. C. & Nishiizumi, K. Beryllium 10 concentrations in the Greenland Ice Sheet Project 2 ice core from 3–40 ka. *J. Geophys. Res.* **102**, 26699–26706 (1997).
- Goslar, T. *et al.* High concentration of atmospheric ^{14}C during the Younger Dryas cold episode. *Nature* **377**, 414–417 (1995).
- Björck, S. *et al.* Synchronized terrestrial-atmospheric deglacial records around the North Atlantic. *Science* **274**, 1155–1160 (1996).
- Stocker, T. F. & Wright, D. G. Rapid changes in ocean circulation and atmospheric radiocarbon. *Paleoceanography* **11**, 773–795 (1996).
- Marchal, O. *et al.* Modelling the concentration of atmospheric CO_2 during the Younger Dryas climate event. *Clim. Dyn.* **15**, 341–354 (1999).
- Stuiver, M. & Polach, H. A. Discussion reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
- Beer, J. *et al.* Information on past solar activity and geomagnetism from ^{10}Be in the Camp Century ice core. *Nature* **331**, 675–679 (1988).
- Masarik, J. & Beer, J. Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J. Geophys. Res.* **104**, 12099–12111 (1999).
- McHargue, L. R. & Damon, P. E. The global beryllium-10 cycle. *Rev. Geophys.* **29**, 141–158 (1991).
- Broecker, W. S., Peteet, D. M. & Rind, D. Does the ocean-atmosphere system have more than one stable mode of operation? *Nature* **315**, 21–26 (1985).
- Wagner, G. Die kosmogenen Radionuklide ^{10}Be und ^{36}Cl im Summit-GRIP-Eisbohrkern. Thesis, ETH Zürich (1998).
- Wagner, G. *et al.* Chlorine-36 evidence for the Mono Lake event in the Summit GRIP ice core. *Earth Planet. Sci. Lett.* **181**, 1–6 (2000).
- Johnsen, S. J., Dansgaard, W. & White, J. W. C. The origin of Arctic precipitation under present and glacial conditions. *Tellus B* **41**, 452–468 (1989).
- Charles, C. D., Rind, D., Jouzel, J., Koster, R. D. & Fairbanks, R. G. Glacial-interglacial changes in moisture sources for Greenland: Influences on the ice core record of climate. *Science* **263**, 508–511 (1994).
- Siegenthaler, U. Uptake of excess CO_2 by an outcrop-diffusion model of the ocean. *J. Geophys. Res.* **88**, 3599–3608 (1983).
- Stuiver, M. *et al.* INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* **40**, 1041–1083 (1998).
- Stuiver, M. & Quay, P. D. Changes in atmospheric carbon-14 attributed to a variable sun. *Science* **207**, 11–19 (1980).
- Björck, S. *et al.* High-resolution analyses of an early Holocene cooling event may imply solar forcing as an important climate trigger. *Geology* (submitted).
- Alley, R. B. *et al.* Visual-stratigraphic dating of the GISP2 ice core: Basis, reproducibility, and application. *J. Geophys. Res.* **102**, 26367–26381 (1997).
- Hughen, K. A., Overpeck, J. T., Peterson, L. C. & Trumbore, S. Rapid climate changes in the tropical Atlantic region during the last deglaciation. *Nature* **380**, 51–54 (1996).
- Tauxe, L. Sedimentary records of relative paleointensity of the geomagnetic field: theory and practice. *Rev. Geophys.* **31**, 319–354 (1993).
- Alley, R. B. *et al.* Changes in continental and sea-salt atmospheric loadings in central Greenland during the most recent deglaciation: model-based estimates. *J. Glaciol.* **41**, 503–514 (1995).
- Johnsen, S. J., Dahl-Jensen, D., Dansgaard, W. & Gundestrup, N. Greenland palaeotemperatures derived from GRIP bore hole temperature and ice core isotope profiles. *Tellus B* **47**, 624–629 (1995).
- Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
- Johnsen, S. J. *et al.* The $\delta^{18}\text{O}$ record along the Greenland Ice Core Project deep ice core and the problem of possible Eemian climatic instability. *J. Geophys. Res.* **102**, 26397–26410 (1997).

30. Groote, P. M. & Stuiver, M. Oxygen 18/16 variability in Greenland snow and ice with 10^{-3} - to 10^5 -year time resolution. *J. Geophys. Res.* **102**, 26455–26470 (1997).

Acknowledgements

Data were provided by the National Snow and Ice Data Center, University of Colorado at Boulder, and the WDC-A for Paleoclimatology, National Geophysical Data Center, Boulder, Colorado. We thank K. Hughen for the ^{14}C calibration data and the grey scale record of the Cariaco sediments. This work was supported by the Swiss National Science Foundation and the US Department of Energy.

Correspondence and requests for materials should be addressed to R.M. (e-mail: muscheler@eawag.ch).

Triggering of earthquake aftershocks by dynamic stresses

Deborah Kilb^{*†}, Joan Gomberg[‡] & Paul Bodin^{*}

^{*} Center for Earthquake Research and Information, The University of Memphis, 3876 Central Avenue, Suite 1, Memphis, Tennessee, 38152, USA

[‡] US Geological Survey, Center for Earthquake Research and Information, 3876 Central Avenue, Suite 2, Memphis, Tennessee, 38152, USA

[†] Present address: Department of Geosciences, Princeton University, Princeton, New Jersey, USA.

It is thought that small 'static' stress changes due to permanent fault displacement can alter the likelihood of, or trigger, earthquakes on nearby faults¹. Many studies of triggering in the near-field, particularly of aftershocks, rely on these static changes as the triggering agent^{2–4} and consider them only in terms of equivalent changes in the applied load on the fault^{3–6}. Here we report a comparison of the aftershock pattern of the moment magnitude

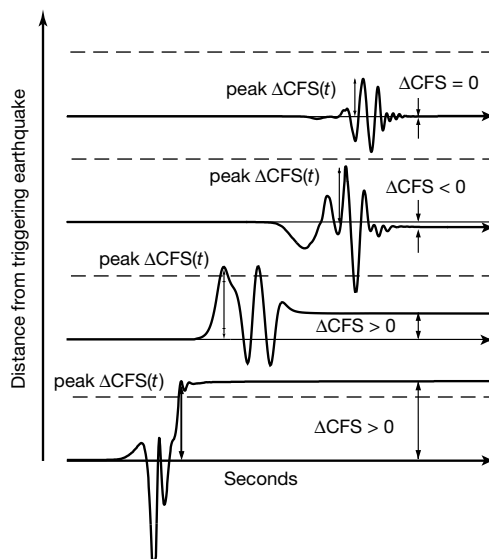


Figure 1 Cartoon time-histories of complete Coulomb stress change, $\Delta\text{CFS}(t)$, and its variation with distance. $\Delta\text{CFS}(t) = \Delta\tau(t) + \mu(\Delta\sigma(t) - B\Delta p(t))$ where $\Delta\tau$ and $\Delta\sigma$ denote changes in shear and normal stresses resolved on a fault plane, μ is the coefficient of friction, Δp is the change in isotropic compressional stress, and B is similar to Skempton's coefficient¹. Double arrows show peak $\Delta\text{CFS}(t)$ and static ΔCFS measurements. Dashed lines indicate hypothetical stress change threshold; peak $\Delta\text{CFS}(t)$ exceeds this threshold at the two closest sites whereas the static Coulomb stress change exceeds it only at the closest site. ΔCFS thresholds are typically several per cent or less of the stress relieved by an earthquake¹.

$M_w = 7.3$ Landers earthquake, not only with static stress changes but also with transient, oscillatory stress changes transmitted as seismic waves (that is, 'dynamic' stresses). Dynamic stresses do not permanently change the applied load and thus can trigger earthquakes only by altering the mechanical state or properties of the fault zone. These dynamically weakened faults may fail after the seismic waves have passed by, and might even cause earthquakes that would not otherwise have occurred. We find similar asymmetries in the aftershock and dynamic stress patterns, the latter being due to rupture propagation, whereas the static stress changes lack this asymmetry. Previous studies have shown that dynamic stresses can promote failure at remote distances^{7–12}, but here we show that they can also do so nearby.

Earthquake-generated stress changes may enhance (or diminish) the tectonic load applied to another fault, moving it closer to (or further from) failure. Most studies have looked at permanent

changes in a linear combination of stress components known as the static Coulomb failure stress change (ΔCFS). ΔCFS may advance (or retard) failure by the approximate time required to accumulate an equivalent load at the tectonic loading rate. These times are often referred to as clock-advances (or clock-delays). Presumably earthquakes triggered by a clock-advance would have taken place at some later date.

Earthquakes also radiate seismic waves, which transmit transient, oscillatory stresses that do not permanently alter the net load on a fault. However, both dynamic and static stress changes may modify the properties of the fault zone itself, effectively changing the failure criteria¹³ and/or the nature and rates of mechanical or chemical processes that lead to earthquake rupture. Changes in the mechanical properties or failure processes of a fault do not mean that failure occurs immediately or with constant clock-advance; thus they allow variable time delays between triggering and triggered earthquakes.

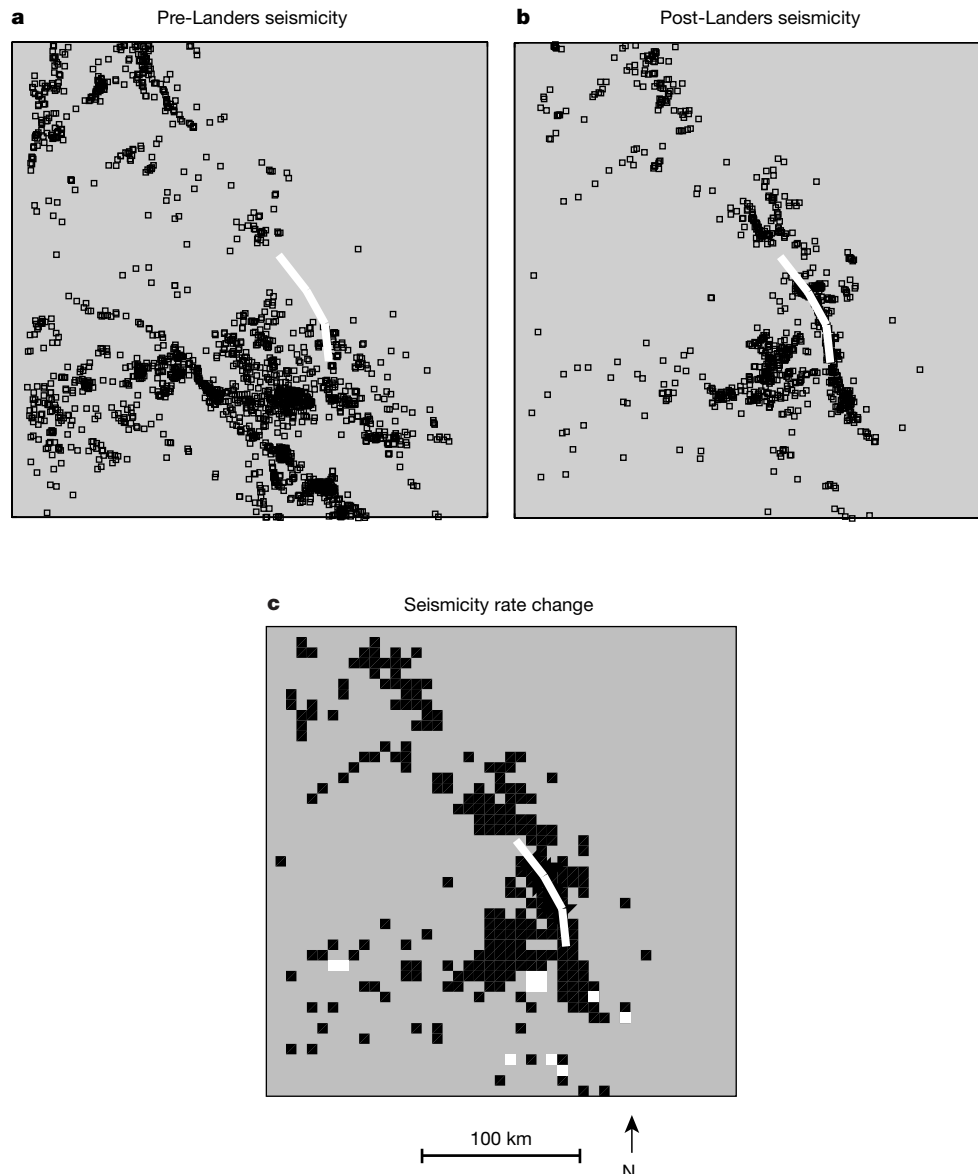


Figure 2 Pre- and post-Landers seismicity and calculated seismicity rate change. The regions of increased seismicity rate are very asymmetrically distributed about the Landers rupture (thick white line), with most of the increases occurring to the northwest. Maps of epicentres (squares) from 7/84 to 4/92 (**a**) and from 6/92 to 3/93 (**b**). This earlier background period is free of magnitude $M > 6$ events, significant network changes, and recording gaps. Catalogue completeness is assured by including only events of $M \geq 2.2$. Use of a declustered catalogue helps satisfy the assumption of independence between

earthquakes (other than with the Landers mainshock)^{24–25}. **c**, Seismicity rate change map derived using a modified β -statistic^{21–22} which measures the difference between the observed number of post-Landers events in a grid cell and the number expected based on an average seismicity rate determined using the entire event population. Normalization of the difference by the standard deviation of the seismicity rate accounts for natural rate fluctuations. Black indicates increased rate, white decreased rate, and grey no rate change.

The delayed failure observed in aftershock sequences has been explained in terms of an approximate rate-state frictional-instability model and triggering by static stress changes². However, this same frictional model driven by dynamic stresses does not predict the delays required to generate an aftershock sequence, perhaps because it assumes fixed constitutive properties and/or that aftershocks are inevitable clock-advanced events. Thus, dynamic stress triggering may imply that aftershocks would never have occurred in the absence of the mainshock, being instead triggered by a change in the state of faults. Despite the difficulty of testing this idea, a few field observations¹² and theoretical models^{14,15} suggest that transient stresses may generate new earthquake-nucleating instabilities.

Transient and static changes are inseparable temporally (Fig. 1) so we refer to the 'complete' Coulomb stress change, denoted as $\Delta\text{CFS}(t)$. We propose that positive (in the direction of tectonic loading) maximum or peak values of $\Delta\text{CFS}(t)$, which occur during the transient stress changes and may include some component of static stress change, should correlate with increases in seismicity rate following a triggering earthquake. Sufficiently large positive stress changes may greatly alter and weaken faults, enhancing their likelihood of failure. Although the specific mechanisms that lead to instability and earthquake nucleation on crustal faults remain conjectural and are not addressed here, it is reasonable to assume that transient shear stresses cause irreversible weakening. For example, strengthening processes such as crack healing are too slow to counteract the effects of mechanisms such as sub-critical crack growth¹⁶. Also, if nucleation occurs as frictional instability governed by commonly assumed rate-state constitutive laws², then a rapid positive stress change weakens a fault much more than the strengthening that may occur owing to a rapid stress change of equal amplitude but opposite sign¹⁵. Laboratory data addressing differences between static and dynamic stresses on fault strength are scant, and one recent experiment shows that normal stress vibrations could lead to weakening or strengthening¹⁷. We employ earthquake data and theoretical calculations of stress changes, and ask what changes in the dynamic and static Coulomb stress are necessary to trigger observed seismicity rate changes. We do not expect peak $\Delta\text{CFS}(t)$ to lead to seismicity decreases.

Permanent positive and negative changes in the loading bring a fault closer to or further from failure, respectively, so we expect that the sign of ΔCFS should correlate with increases (positive) and decreases (negative) in seismicity rates. ΔCFS may also alter and weaken faults, but we discount this possibility because of its relatively much smaller amplitude (see below). We expect any correlation between $\Delta\text{CFS}(t)$ and seismicity rate changes to be imperfect, because other processes and factors undoubtedly also influence seismicity rates⁹. The Landers earthquake sequence in California is ideal for testing these hypotheses because the mainshock has a relatively simple and well-constrained source mechanism^{18,19}, and took place in a relatively well-instrumented region²⁰. One study of the Landers earthquake documents a strong correlation between ΔCFS and the spatial distribution of its aftershocks⁴. A similar correlation might reasonably exist with peak $\Delta\text{CFS}(t)$, because it usually exceeds ΔCFS . Beyond about 10 km from the causative rupture, peak $\Delta\text{CFS}(t)$ is typically more than an order of magnitude larger than ΔCFS .

We derive a map of seismicity rate change attributable to the Landers earthquake using the Southern California Seismic Network (SCSN) earthquake catalogue. We spatially grid the catalogued epicentres and compute the seismicity rate change in each grid cell using a modified version of the β -statistic^{21,22}. We assume a rate change is significant when $|\beta| \geq 1$ which, for a Poisson process, would be significant only at the 65% level. This choice of criterion compensates somewhat for the fact that seismicity is not poissonian, and for the effect of gridding the area, which causes the estimated rate change to appear more variable than both our intuition and the $\Delta\text{CFS}(t)$ maps suggest is appropriate. To avoid making unwarranted assumptions about the magnitude of the rate change and its relationship to $\Delta\text{CFS}(t)$, we assign rate increases, no rate change, and rate decreases values of 1, 0, and -1, respectively (Fig. 2).

We calculate $\Delta\text{CFS}(t)$ using a discrete-wavenumber reflectivity program²³. We constrain model parameters by modelling the Landers' mainshock seismic displacements recorded at nearby broadband TERRAscope stations, and assuming that parameters that control surface displacements also control $\Delta\text{CFS}(t)$ at depth. The correlation between maps of Coulomb stress change and

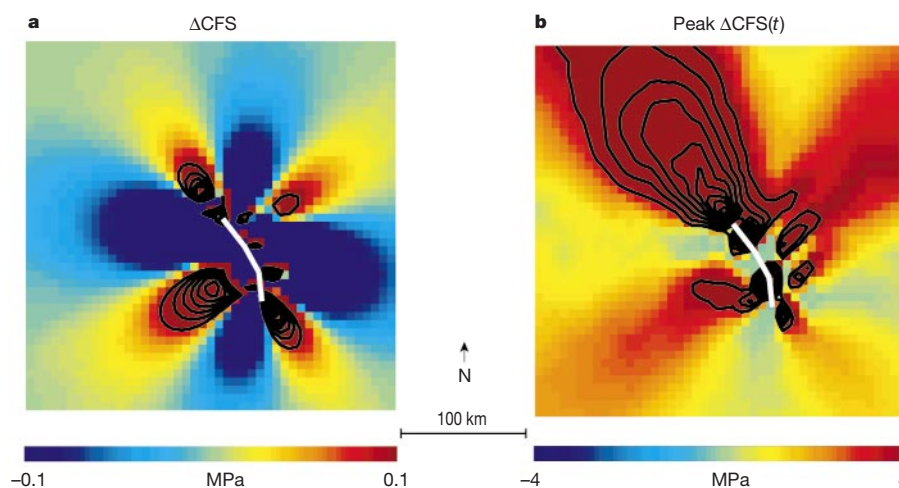


Figure 3 Maps of modelled ΔCFS and peak $\Delta\text{CFS}(t)$. Notable differences between these maps are the greater asymmetry about the Landers rupture (white lines) and much larger amplitudes in the peak $\Delta\text{CFS}(t)$ map. The amplitude scales are chosen to optimize the correlation with the seismicity rate change (see text). Calculations require resolving the stress field on the fault planes of the triggered events. We chose mean values of strongly peaked histograms of aftershock focal mechanism parameters and depths reported in the SCSN catalogue as the most appropriate (strike, 330° ; dip, 90° ; rake, 180° ; depth, 4.5 km). This fault plane is also nearly optimally oriented given the regional stress field and approximately parallels the general strike of mapped faults. Positive $\Delta\text{CFS}(t)$ is that which would produce slip in the observed rake direction. We assume Coulomb parameters (see

Fig. 1 legend) of $\mu = 0.6$ and $B = 0.85$, in accord with previous studies of ΔCFS ¹.

a, ΔCFS map contoured at intervals of 0.05 MPa for values ≥ 0.1 MPa and shaded for smaller values. The addition of the $M = 6.1$ Joshua Tree and $M = 6.5$ Big Bear earthquakes to the stress calculations causes the high positive stress lobes in the southwest and southeast to become slightly more heterogeneous (that is, mixed with small regions of negative stress change). **b**, Peak $\Delta\text{CFS}(t)$ map contoured at intervals of 1.5 MPa for values ≥ 4 MPa and shaded for smaller values. Addition of the Joshua Tree and Big Bear earthquakes causes the high in the southwest to elongate southwesterly by a few tens of kilometres.

seismicity rate change depends on the range of stress change values considered significant. For peak $\Delta\text{CFS}(t)$ and for $|\Delta\text{CFS}|$ we search for the minimum value, or threshold, that optimizes the correlation (that is, the smallest contour levels in Fig. 3a and b). We interpret this as constraining the stress change threshold (assumed here to be regionally uniform) required to alter the fault rheology or the load enough to change the failure rate. For a given threshold, if the extreme values of peak $\Delta\text{CFS}(t)$ or ΔCFS in each cell exceed the threshold the cell is assigned a value of 1, if the extreme value is below the negative of the threshold it is assigned a value of -1 , and all others are assigned zeros.

We measure the correlation by counting the number of grid cells in which both the Coulomb stress and seismicity rate changes match. Allowing for the possibility that a rate decrease has gone undetected in regions of low rates, owing to our short observation period, we also increment the count for grid cells in which ΔCFS equals -1 and the seismicity rate change is zero as a result of no pre- and post-Landers seismicity. The significance of the correlation count for each threshold tested is estimated by comparing it to the statistical distribution of counts obtained by repeating the procedure 500 times, using spatially randomized seismicity rate change

maps. Significant counts, or correlations, must differ from the mean of the randomized counts by two or more standard deviations and the optimal threshold produces the most significant correlation. The optimal triggering thresholds at a depth of 4.5 km, corresponding to the mean aftershock depth, equal 0.1 MPa for ΔCFS (Fig. 3a) and 4.0 MPa for peak $\Delta\text{CFS}(t)$ (Fig. 3b). This static threshold is consistent with that found in other studies^{1,3}. Significant correlations exist for thresholds of 0.001 to 0.5 MPa for ΔCFS and more than 0.5 MPa for peak $\Delta\text{CFS}(t)$. We also computed maps of ΔCFS and peak $\Delta\text{CFS}(t)$ at the surface, and at depths of 2 km and 11 km. Although the absolute amplitudes of ΔCFS and peak $\Delta\text{CFS}(t)$ decrease slightly with depth, the mapped patterns do not change significantly. Correlation results are similarly robust with respect to plausible variations in the assumed geometry of aftershock fault planes.

Both ΔCFS and peak $\Delta\text{CFS}(t)$ correlate statistically with seismicity rate change with a high degree of confidence (well outside the tails of the distribution of random correlations). Despite this high confidence, our statistical method suggests that both ΔCFS and peak $\Delta\text{CFS}(t)$ predict seismicity rate changes no more than 7% better than a random process. However, a qualitative visual assessment of the correlations (Fig. 4) indicates that this is clearly overly conservative, because the eye smooths the seismicity rate change maps more in accord with the variability of the stress changes than does our statistical measure. Because earthquakes are represented as point data, the β -statistic depends strongly on grid cell size. The $6 \times 6 \text{ km}^2$ cell size was chosen to capture accurately the fluctuation in stress changes close to the mainshock fault, and whereas the statistical correlations could be enhanced if we increased our cell size or applied a smoothing operator, it is not clear how to do this objectively.

Statistically ΔCFS and peak $\Delta\text{CFS}(t)$ correlate with seismicity rate change equally well. At the closest distances, only an amplitude scaling relationship differs between the two. Farther out the asymmetry in the aftershock distribution correlates better with the asymmetry in the peak $\Delta\text{CFS}(t)$ map, which is absent in the ΔCFS map. The peak $\Delta\text{CFS}(t)$ asymmetry results from the mainshock's unilateral northward rupture propagation, and its correlation with the aftershock pattern leaves little doubt that the peak stresses, although transient, exert a controlling influence. Even our more conservative statistical results imply that the many previous studies that suggest a causal relationship between ΔCFS and seismicity rate change^{3,22} may be missing an important part of the triggering process. Moreover, we expect dynamic stresses, thanks to their much larger amplitude relative to the static changes, to alter fault zone properties more effectively in ways that permit a range of time-delays between triggering and triggered events. Permanent stress changes can also contribute to the alteration of fault zone properties, but our results indicate that this effect may be relevant only extremely close to the mainshock rupture where the permanent stress changes may be nearly as large as the dynamic stresses. Preliminary analyses of the moment magnitude $M_w = 7.1$ Hector Mine, California earthquake of 16 October 1999 also show a strong correlation between its rupture dynamics (directivity direction) and the change in seismicity rate at remote distances. Although the existence of a similar correlation in the near-field remains to be determined, these preliminary results suggest that nature may have just reconfirmed the lessons offered by the Landers earthquake. □

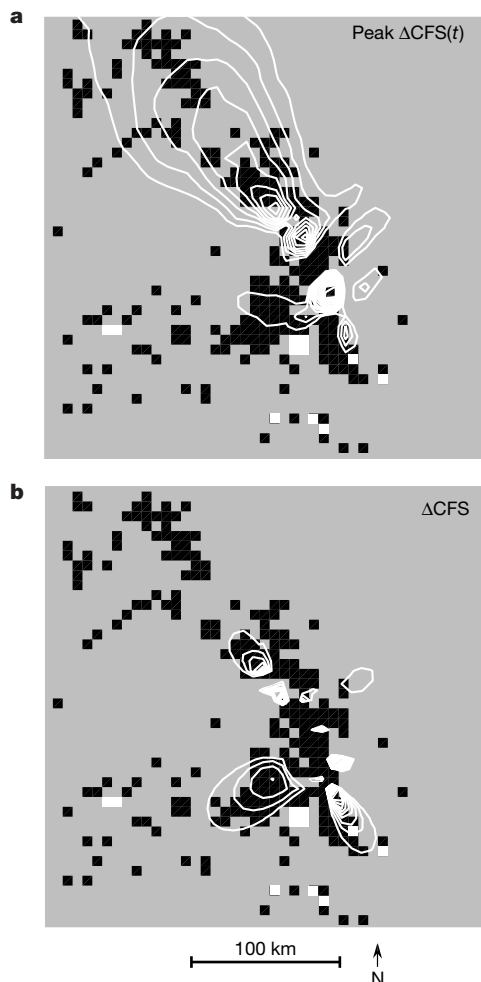


Figure 4 Comparisons of peak $\Delta\text{CFS}(t)$ and ΔCFS with seismicity rate change. The peak $\Delta\text{CFS}(t)$ contours and the pattern of seismicity rate change share the same asymmetry, which is lacking in the ΔCFS contours. Contours (from Fig. 3) show (a) peak $\Delta\text{CFS}(t) \geq 4 \text{ MPa}$ (interval, 1.5 MPa) and (b) $\Delta\text{CFS} \geq 0.1 \text{ MPa}$ (interval, 0.05 MPa). Map shading indicates significant rate increases (black), decreases (white) and regions of no change (grey). The correlation of seismicity rate change with peak $\Delta\text{CFS}(t)$ improves slightly with the addition of the $M = 6.1$ Joshua Tree and $M = 6.5$ Big Bear earthquakes to the stress calculations, and has negligible effect on the correlation with ΔCFS .

Received 10 January; accepted 28 August 2000.

- Harris, R. A. Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *J. Geophys. Res.* **103**, 24347–24358 (1998).
- Dieterich, J. A. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99**, 2601–2618 (1994).
- Toda, S., Stein, R. S., Reasenberg, P. A., Dieterich, J. & Yoshida, A. Stress transferred by the 1995 Mw=6.9 Kobe, Japan shock: Effect on aftershocks and future earthquake probabilities. *J. Geophys. Res.* **103**, 24543–24566 (1998).
- King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).

5. Jaume, S. C. & Sykes, L. R. Changes in state of stress on the southern San Andreas Fault resulting from the California earthquake sequence of April to June, 1992. *Science* **258**, 1325–1328 (1992).
6. Harris, R. A. & Simpson, R. W. Changes in static stress on southern California faults after the 1992 Landers earthquake. *Nature* **360**, 251–254 (1992).
7. Anderson, J. G. *et al.* Seismicity in the western Great Basin apparently triggered by the Landers, California, earthquake, 28 June 1992. *Bull. Seismol. Soc. Am.* **84**, 863–891 (1994).
8. Gombert, J. & Bodin, P. Triggering of the Little Skull Mountain, Nevada earthquake with dynamic strains. *Bull. Seismol. Soc. Am.* **84**, 844–853 (1994).
9. Gombert, J. Stress/strain changes and triggered seismicity following the Mw7.3 Landers, California, earthquake. *J. Geophys. Res.* **101**, 751–764 (1996).
10. Gombert, J. & Davis, S. Stress/strain changes and triggered seismicity at The Geysers, California. *J. Geophys. Res.* **101**, 733–749 (1996).
11. Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
12. Hill, D. P., Johnston, M. J. S., Langbein, J. O. & Bilham, R. Response of Long Valley caldera to the Mw=7.3 Landers, California, earthquake. *J. Geophys. Res.* **100**, 12985–13005 (1995).
13. Gombert, J., Beeler, N. M., Blanpied, M. L. & Bodin, P. Earthquake triggering by transient and static deformations. *J. Geophys. Res.* **103**, 24411–24426 (1998).
14. Boatwright, J. & Cocco, M. Frictional constraints on crustal faulting. *J. Geophys. Res.* **101**, 13895–13909 (1996).
15. Gombert, J., Blanpied, M. L. & Beeler, N. M. Transient triggering of near and distant earthquakes. *Bull. Seismol. Soc. Am.* **87**, 294–309 (1997).
16. Atkinson, B. K. Subcritical crack growth in geologic materials. *J. Geophys. Res.* **89**, 4077–4114 (1984).
17. Richardson, E. & Marone, C. Effects of normal stress vibrations on frictional healing. *J. Geophys. Res.* **104**, 28857–28878 (1999).
18. Dreger, D. S. Investigation of the rupture process of the 28 June 1992 Landers earthquake utilizing TERRASCOPE. *Bull. Seismol. Soc. Am.* **84**, 713–724 (1994).
19. Wald, D. J. & Heaton, T. H. Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake. *Bull. Seismol. Soc. Am.* **84**, 668–691 (1994).
20. Hauksson, E., Jones, L. M., Hutton, K. & Eberhart-Phillips, D. The 1992 Landers earthquake sequence: seismological observations. *J. Geophys. Res.* **98**, 19835–19858 (1993).
21. Matthews, M. V. & Reasenberg, P. A. Statistical methods for investigating quiescence and other temporal seismicity patterns. *Pure Appl. Geophys.* **126**, 357–372 (1988).
22. Reasenberg, P. A. & Simpson, R. W. Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science* **255**, 1687–1690 (1992).
23. Cotton, F. & Coutant, O. Dynamic stress variations due to shear faults in a plane-layered medium. *Geophys. J. Int.* **128**, 676–688 (1997).
24. Frohlich, C. & Davis, S. D. Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogs. *Geophys. J. Int.* **100**, 19–32 (1990).
25. Davis, S. D. & Frohlich, C. Single-link cluster analysis of earthquake aftershocks: decay laws and regional variations. *J. Geophys. Res.* **96**, 6335–6350 (1991).

Acknowledgements

We are grateful to F. Cotton, O. Coutant, P. Reasenberg and S. Davis for kindly supplying us with their computer codes. We thank S. Hough, M. Johnston & C. Marone for thoughtful reviews. The US Geological Survey funded this work.

Correspondence and requests for materials should be addressed to J.G.
(e-mail: gombert@usgs.gov).

Geochemical evidence for terrestrial ecosystems 2.6 billion years ago

Yumiko Watanabe^{*†}, Jacques E. J. Martini[‡] & Hiroshi Ohmoto^{*}

^{*} Astrobiology Research Center and Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

[†] Faculty of Science, Tohoku University, Sendai, 980-8578, Japan

[‡] Geological Survey of South Africa, Private Bag X112, Pretoria, South Africa

Microorganisms have flourished in the oceans since at least 3.8 billion years (3.8 Gyr) ago^{1,2}, but it is not at present clear when they first colonized the land. Organic matter in some Au/U-rich conglomerates and ancient soils of 2.3–2.7 Gyr age has been suggested as remnants of terrestrial organisms^{3–5}. Some 2.7-Gyr-old stromatolites have also been suggested as structures created by terrestrial organisms^{6–7}. However, it has been disputed whether this organic matter is indigenous or exogenic, and whether these stromatolites formed in marine or fresh water. Consequently, the oldest undisputed remnants of terrestrial organisms are currently the 1.2-Gyr-old microfossils from Arizona, USA⁸. Unusually carbonaceous ancient soils—palaeosols—have been found in

the Mpumalanga Province (Eastern Transvaal) of South Africa⁹. Here we report the occurrences, elemental ratios (C, H, N, P) and isotopic compositions of this organic matter and its host rocks. These data show that the organic matter very probably represents remnants of microbial mats that developed on the soil surface between 2.6 and 2.7 Gyr ago. This places the development of terrestrial biomass more than 1.4 billion years earlier than previously reported.

In the Mpumalanga Province of South Africa, many palaeosols have been recognized on the Archaean basement complex, which is composed primarily of granites, granite-gneisses, basalts and minor ultramafic rocks (for example, dunite, peridotite and serpentinite) of age 2.7–3.6 Gyr. These palaeosols are overlain by various formations of the Transvaal Supergroup (2.1–2.6 Gyr), and were subjected to the zeolite- to lower green-schist facies metamorphism (temperature $T < \sim 300^\circ\text{C}$) associated with the intrusion of the 2.05-Gyr Bushveld Igneous Complex (Fig. 1) (ref. 9). Palaeosols on granites and basalts often exceed 20 m in thickness. They are easily recognizable because of the abundant aluminosilicates (feldspars, biotite, hornblende) in the parental rocks that were converted to less-soluble, pale, Al_2O_3 -rich clays, such as kaolinite, smectite and sericite. In contrast, thick palaeosols rarely develop on ultramafic rocks because such rocks are mostly composed of aluminium-poor silicates (olivine, pyroxene) that are easily leached out by rain water. Exceptionally thick soils may develop on ultramafic rocks when the porous spaces created by dissolution of silicate minerals are continuously filled by aerosols or by precipitation of new minerals (for example, calcite, dolomite and silica) from soil water and/or ground water. Such palaeosols were discovered at Schagen, Elandshoek and Kalkkloof areas in Mpumalanga Province (Fig. 1) (ref. 9). The data we give here are from the organic matter at Schagen.

The palaeosol section at Schagen developed on a >2.7-Gyr serpentinitized dunite, and is overlain by a quartzite bed (~30-m thick) of the 2.6-Gyr Black Reef Formation. Therefore, the palaeosol

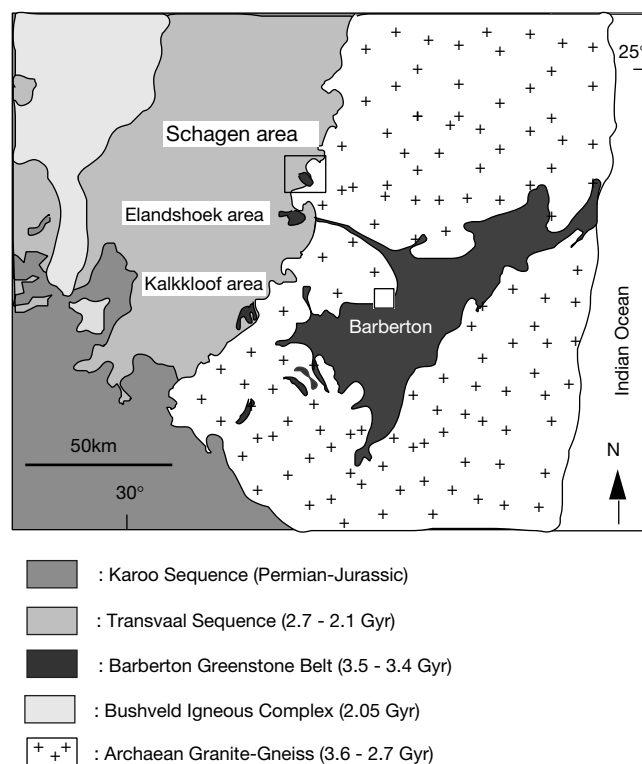


Figure 1 A simplified map of Mpumalanga Province, South Africa. The figure shows the main geological units, and the localities of organic-rich palaeosols (Schagen, Elandshoek, Kalkkloof) (modified after ref. 9).

probably developed between 2.6 and 2.7 Gyr ago. It is about 17 m thick, and is characterized by the abundance of calcite (CaCO_3) and dolomite ($\text{CaMg}(\text{CO}_3)_2$), especially in the upper ~ 7 m section, and of quartz grains in the uppermost ~ 0.7 m section (see Fig. 2 and Methods). This soil section seems to have developed under semi-arid conditions through the following three processes: (1) dissolution by local rain and ground water of silicates in the parental rock (mostly serpentinized dunite) and of carbonates formed during soil formation; (2) precipitation of calcite and dolomite in the pores and also on the surface of soil by evaporation of locally discharging ground water; and (3) accumulation on the soil surface of aerosols (mostly quartz and aluminous clays) from the neighbouring granite terrain. Process (1) mostly occurred during rainy seasons, and (2) and (3) mostly during dry seasons. These processes are essentially identical to those proposed for the formation of modern calcite-rich soils (calcrete) and dolomite-rich soils (dolocrete)¹⁰. Process (2) was necessary to explain the abundance of Ca (up to 54% CaO) in the rock that originally contained only ~ 0.1 wt% CaO. The ground water probably obtained Ca^{2+} from the nearby granites (~ 3 wt% CaO). The Mg in the dolomite of the palaeosol was probably derived from the local Mg-rich serpentinite. Most of the groundwater HCO_3^- that formed carbonates probably originated from CO_2 in the Archaean atmosphere.

The content of reduced carbon in 32 bulk-rock samples (each $\sim 30 \text{ cm}^3$ in volume) varies from less than 0.01 wt% (the detection limit) to 0.36 wt% (Fig. 2). Samples with > 0.1 wt% reduced C are mostly restricted to the upper ~ 5 m section, from zones I–IV. In comparison, the reduced-C content of a typical Precambrian palaeosol is less than 0.1 wt% (ref. 11). Therefore, the important questions are the origin(s) and the concentration mechanism(s) of reduced C in the Schagen palaeosol, whether the reduced C was formed by biological or abiological process(es), and whether it accumulated before, during or after soil formation. We have evaluated the following possibilities for the origin of the reduced

C: first, it represents disseminated graphite crystals that formed > 2.7 Gyr ago during serpentinization of dunite at $T > 400^\circ\text{C}$ and subsequently concentrated during soil formation sometime between 2.6 and 2.7 Gyr ago; second, it represents remnants of liquid hydrocarbons (for example, petroleum) introduced after soil formation; and third, it represents remnants of biomats developed on the soil surface. The first possibility needs to be evaluated because the association of ultramafic rocks and non-biogenic graphite is not uncommon¹². The second possibility needs to be evaluated because migration of petroleum is a common phenomenon. The third possibility is, however, strongly supported by the texture, crystallinity, abundance ratios of bio-essential elements, and most importantly, the mode of occurrence of the organic matter in these rocks (see below).

The crystallinity and H/C ratio of the organic matter (see Methods) suggest that it is not of igneous or hydrothermal origin, but that it was incorporated in the soil as a biogenic product before the 2.0-Gyr metamorphic event. The existence of positive correlations among the contents of bio-essential elements (carbon, nitrogen and phosphate) is one of several lines of evidence that the organic matter is an *in situ* remnant of ~ 2.6 -Gyr-old organisms, rather than of liquid hydrocarbons (see Fig. 3 and Methods).

If the organic matter was introduced in the palaeosol after soil formation, some, if not most, organic matter would be expected to occur along fractures in the rocks. But this is not the case: instead, the organic matter is almost always concentrated in clay-rich parts of the rocks (Fig. 4a–d). Organic matter and clays (talc, chlorite, ferristilpnomelane) are so intimately mixed together that the size and morphology of individual ‘grains’ of organic matter can only be recognized under the electron microscope (Fig. 4c and d). In the uppermost soil zone (zone I), which is mostly composed of aerosols, the organic-rich clays are mostly ferristilpnomelane with minor chlorite. (These Fe-bearing aluminosilicates probably formed during regional metamorphism from precursor soil minerals,

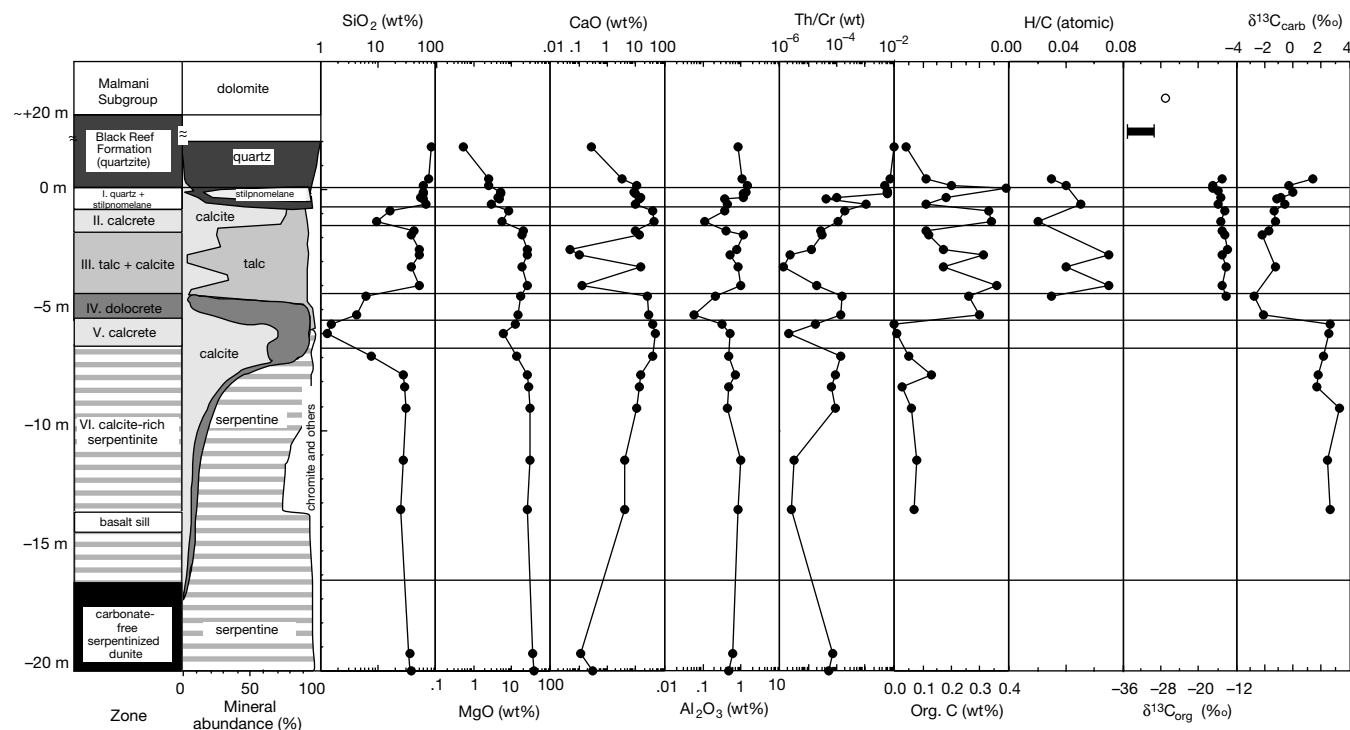


Figure 2 Mineralogical and chemical characteristics of the 17-m palaeosol section at Schagen (see Methods for details). The parental rock is carbonate-free serpentinite. The uppermost palaeosol (zone I) is overlain by the 2.6-Gyr Black Reef Formation. The solid bar in the $\delta^{13}\text{C}_{\text{org}}$ graph represents the $\delta^{13}\text{C}$ values of organic matter in shales from the

Black Reef Formation at other localities; the open circle represents the $\delta^{13}\text{C}$ value of organic matter in a stromatolite from the Malmani subgroup at Schagen. $\delta^{13}\text{C} = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{standard}}] - 1$; where the standard is PDB (Belemnite from the Pee Dee Formation).

such as smectite and goethite). They typically occur as seams ($\sim 30\text{ }\mu\text{m}$ to $\sim 1\text{-mm}$ thick, and $> 1\text{-cm}$ long) between fine-grained and coarse-grained layers of quartz in rocks exhibiting sorted-grain textures (Fig. 4a), as networks surrounding minor quartz grains in ferristilpnomelane-rich rocks, and as clots ($0.1\text{--}5\text{ mm}$ diameter) and disseminated grains ($< 0.1\text{ mm}$) in quartz-rich rocks (Fig. 4d). These features suggest that the organic matter in the uppermost soil zone is an indigenous remnant of microbial mats that developed on the surface of clay-rich soil during the periods of low fluxes of aerosols (that is, the rainy seasons); the mats were blanketed by aerosols deposited mostly during the dry seasons. The original

thickness of a microbial mat, estimated from the crystallinity and the H/C ratio of the remnant, was probably more than 10 times the thickness of the remnant; thus it was probably more than about $0.3\text{--}1\text{-cm}$ thick.

Organic-rich clays in the lower palaeosol sections (zones II–IV) are mostly talc and minor chlorite. They typically occur as layers ($0.1\text{--}3\text{-mm}$ thick) along the irregular boundaries between different carbonate mineral generations in rocks with ‘stylolite’ textures (Fig. 4b). Such features may suggest reconcentration, through dissolution/precipitation processes of carbonate, of organic matter and clays. The organic matter in these soil zones may

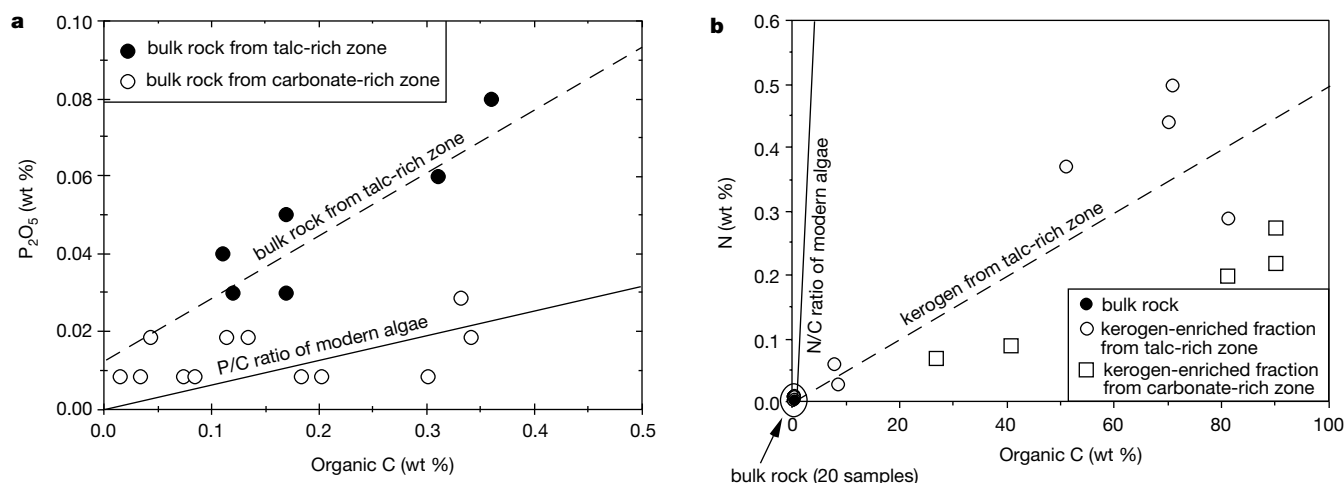


Figure 3 Concentration relationships between organic carbon and phosphorus, and organic carbon and nitrogen. **a**, Organic-carbon/phosphorus relationship in bulk-rock samples; **b**, organic-carbon/nitrogen relationship in bulk-rock samples and in kerogen-

enriched fractions. These latter fractions are what remains after the carbonates, silicates and oxides in the rock samples have been dissolved by hydrochloric and hydrofluoric acids.

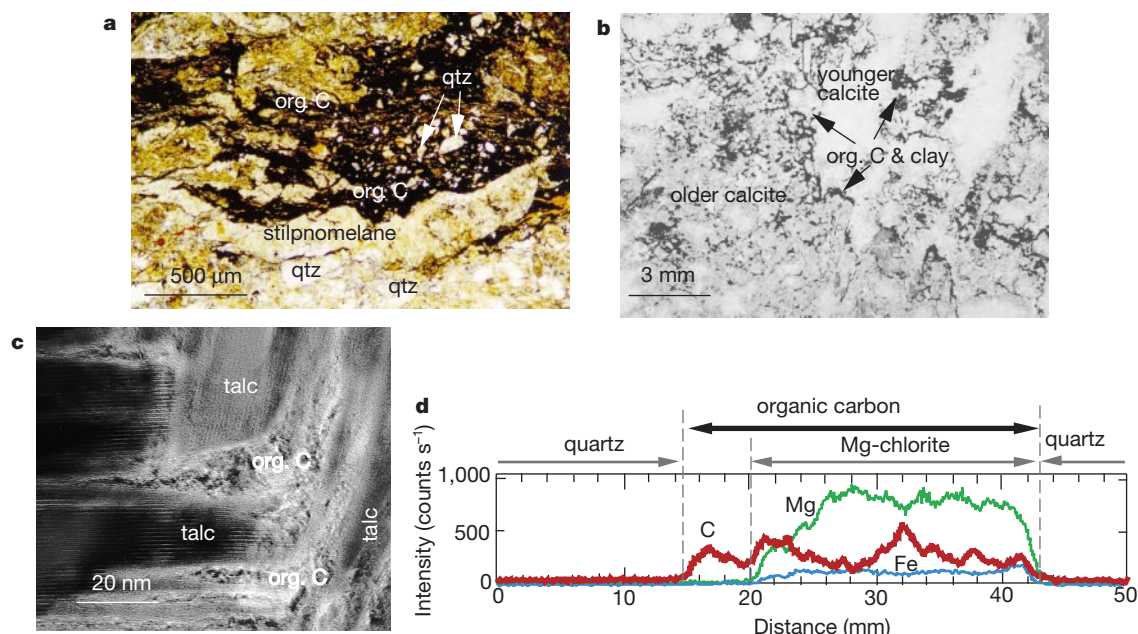


Figure 4 Association of organic matter with clays in the palaeosol. **a**, Thin-section photograph of a sample from zone I. Seams of organic matter (black) typically occur in ferristilpnomelane-rich areas (light brown) in rock that is mostly composed of graded beds of aerosol quartz with minor calcite. **b**, Thin-section photograph of a calcrite sample from zone II. At least two generations of calcite are recognized: earlier calcite, rich in talc and organic C (grey); and later calcite, free of talc and organic C (white). This indicates repeated events of dissolution/precipitation of calcite. Seams of organic-C-rich clays, probably fossilized microbial mats, appear to have been distorted during the dissolution/

precipitation events of calcite. **c**, A transmission electron microscope image of a portion of an organic-rich clay layer in the same sample of **b**. The black area is well-crystallized talc with a $9.45\text{-}\text{\AA}$ lattice parameter. The white area with dark masses is amorphous graphite (organic matter) with a discontinuous and wavy layer structure. **d**, Concentration profiles of C, Mg and Fe, determined by an electron microprobe analyser, of a sample from the lowest part of zone I. Organic carbon occurs between quartz and chlorite grain boundaries and also within a chlorite-rich layer.

represent fragments of the surface microbial mats that were transported downward by rain water. The lower $\delta^{13}\text{C}$ values (see Fig. 2 legend) for the carbonates of organic-C-rich zones I–IV, compared with those of organic-C-poor zones V and VI (about -3‰ versus about $+2\text{‰}$; see Fig. 2), suggest a partial decomposition of the organic matter during dissolution of carbonates and incorporation of the organically derived HCO_3^- in the formation of new carbonates.

$\delta^{13}\text{C}$ values for organic matter (kerogen) extracted from the upper-four soil zones fall in a narrow range, -17.4‰ to -14.3‰ (Fig. 2). In contrast, the $\delta^{13}\text{C}$ values of organic matter in marine shales of the Black Reef Formation at other localities in Free State Province (Central Transvaal) range from -35.5‰ to -29.7‰ (ref. 13). The $\delta^{13}\text{C}$ value of organic matter in a stromatolite sample of the Malmani Subgroup (marine sediments) at Schagen is -27.7‰ (Fig. 2). The change in $\delta^{13}\text{C}$ of organic matter during metamorphism is only a few per mil (refs 13, 14). Thus, the distinct difference in $\delta^{13}\text{C}$ value between organic matter in the palaeosols and that in the overlying marine sedimentary rocks suggests distinct differences in environments and/or species of the two microbial communities.

Most, if not all, of the organisms comprising the soil surface microbial mats at Schagen between 2.6 and 2.7 Gyr ago were probably photosynthetic organisms. Without molecular evidence, it is difficult to identify the exact types of organisms. The possibility of photosynthetic sulphur bacteria may be ruled out because the environment was essentially sulphur-free (non-marine), as indicated by the virtual absence (that is, below the detection limit of 0.01 wt%) of sulphur in all the palaeosol samples. However, the important role of cyanobacteria in the mat formation at Schagen may be suggested because the carbon isotopic fractionation factor between the Archaean atmospheric CO_2 and the mat organisms, calculated to be about 12‰ (see Methods), is essentially identical to that for modern cyanobacterial mats in fresh water¹⁵. This suggestion is consistent with the recent discovery of molecular fossils of cyanobacteria in 2.7-Gyr-old marine shales¹⁶. Although terrestrial bacterial communities in the Archaean soil were predicted by some previous researchers^{17–19}, this is (to our knowledge) the first study presenting several lines of evidence for an extensive development of microbial mats on the soil surface in the Archaean. Our finding may then imply that an ozone shield developed in the atmosphere more than 2.6 Gyr ago. □

Methods

Mineralogical characteristics of the palaeosol

Serpentinization of the parental dunite probably occurred before the soil formation because serpentinized olivine crystals are replaced by calcite. In the upper ~ 7 m section (zones I–V), essentially all the original silicate minerals in the serpentinite are replaced by calcite, dolomite and/or talc; however, the relic textures of serpentinized olivine are still recognizable in zones III and VI. The degree of replacement then decreases downward in zone VI. Also, euhedral (not rounded) grains of chromite (FeCrO_3), a characteristic accessory mineral of ultramafic rocks, are present in zones I–VI. The above features suggest that zones I–VI were mostly formed by *in situ* alteration of the parental serpentinite dunite.

Quartz is not a primary mineral in the serpentinite, but quartz grains are abundant in the ~ 0.7 -m-thick zone I (Fig. 4a) as well as Fe- and Al-rich clays (chlorite and ferrihydrite). The quartz grains are probably the aerosols from the neighbouring granitic terrain because they are variable in size ($<10\text{ }\mu\text{m}$ to $\sim 100\text{ }\mu\text{m}$ in diameter), irregular in shape, and often contain inclusions of zircon and fluid inclusions of high-temperature ($> 400^\circ\text{C}$) origin. The abundance of aluminous clays and the higher Th/Cr ratios ($\sim 10^{-2}$) for the bulk rocks of zone I, compared with those of the lower zones ($\sim 10^{-3}$), also support the interpretation that zone I contains significant amounts of aerosols from a granitic terrain. In the overlying Black Reef quartzite, the quartz grains are larger, rounder in shape, and uniform in size (0.1–0.6 mm in diameter). As a result, they are probably detrital minerals transported by streams and deposited as beach sands. The Black Reef quartzite is then overlain by the stromatolite-rich Malmani dolomite that probably deposited in a shallow, evaporitic marine environment.

Among the carbonates in massive carbonate zones (zones IV and V), dolomite is generally later than calcite in the paragenesis. These massive carbonate zones may represent the groundwater table during soil formation. In the upper zones (zones II–IV), calcite-rich rocks commonly exhibit textures of a repeated history of dissolution/precipitation, which resemble the stylolite textures (Fig. 4d). Talc ($\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$),

which is abundant in these zones above the massive carbonate zones, was probably formed during the regional metamorphism 2.05 Gyr ago. The precursor soil minerals may have been dolomite and/or Mg-rich clays, such as sepiolite and palygorskite.

Structure and composition of the organic matter

Graphite formed by igneous and/or hydrothermal processes generally has well defined crystal structure and an essentially pure composition (that is, the atomic H/C ratio is <0.01)²⁰. The H/C ratio typically falls between 1 and 1.5 in fresh organisms, but it decreases with the increasing degree of maturation (alteration) of organic matter. According to structure analyses by X-ray diffraction (XRD) and high-resolution transmission electron microscopy (HRTEM), organic matter in the Schagen palaeosol is mostly amorphous, but in some places it is small crystallized grains ($<1\text{ }\mu\text{m}$) of graphite (Fig. 4c). H/C ratios of organic matter (kerogen) from the palaeosol are ~ 0.05 (Fig. 2), which are essentially identical to those of organic matter in typical marine shales of Archaean age that had been subjected to green-schist facies metamorphism²¹.

Nitrogen contents of the bulk-rock samples are all below the detection limit of 0.01 wt%; however, N contents of the extracted kerogen-enriched fractions are much higher, ~ 0.1 – 0.5 wt%. Thus, a positive correlation exists between the N and organic-C contents (Fig. 3b). Another correlation is found between the P_2O_5 and organic-C contents of bulk-rock samples, especially among rocks of the talc-rich zone (zone IV) (Fig. 3a). The Schagen organic matter is lower in N/C, but higher in P/C, ratios compared to the organic matter in modern soils and sediments, presumably because of the diagenetic and metamorphic effects²².

Carbon isotope fractionation

Considering the diagenetic/metamorphic modification^{13,14}, the original $\delta^{13}\text{C}$ values of organic matter in the Schagen palaeosols are estimated to be around -18‰ . The organic-C-poor carbonates in zones I and V ($\delta^{13}\text{C} \approx +2\text{‰}$) that formed in equilibrium with the Archaean atmospheric CO_2 had a $\delta^{13}\text{C}$ value of CO_2 of about -6‰ (ref. 23). The difference in $\delta^{13}\text{C}$ between CO_2 and organic matter was around 12‰, which is within the range for modern cyanobacteria-dominated microbial mats in fresh water (typically between 10 and 15)‰¹⁵.

Received 9 June; accepted 25 September 2000.

1. Mojzsis, S. J. *et al.* Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).
2. Rosing, M. T. ^{13}C -depleted carbon microparticles in >3700 -Ma sea-floor sedimentary rocks from west Greenland. *Science* **283**, 674–676 (1999).
3. Hallbauer, D. K. & van Warmelo, K. T. Fossilized plants in thucholite from Precambrian rocks of the Witwatersrand, South Africa. *Precamb. Res.* **1**, 199–212 (1974).
4. Gay, A. L. & Grandstaff, D. E. Chemistry and mineralogy of Precambrian paleosols at Elliot Lake, Ontario, Canada. *Precamb. Res.* **12**, 349–373 (1980).
5. Mossman, D. J. & Fallow, C. E. G. in *Early Organic Evolution: Implications for Mineral and Energy Resources* (eds Schidlowski, M. *et al.*) 67–75 (Springer, Berlin, 1992).
6. Buck, S. G. Stromatolite and ooid deposits within the fluvial and lacustrine sediments of the Precambrian Ventersdorp supergroup of South Africa. *Precamb. Res.* **12**, 311–330 (1980).
7. Buick, R. The antiquity of oxygen photosynthesis: Evidence from stromatolites in sulfate-deficient Archean lakes. *Science* **255**, 74–77 (1992).
8. Horodyski, R. J. & Knauth, L. P. Life on land in the Precambrian. *Science* **263**, 494–498 (1994).
9. Martini, J. E. J. A Late Archaean–Paleoproterozoic (2.6 Ga) paleosol on ultramafics in the Eastern Transvaal, South Africa. *Precamb. Res.* **67**, 159–180 (1994).
10. Wright, V. P. in *Quantitative Diagenesis: Recent Developments and Applications to Reservoir Geology* (eds Parker, A. & Sellwood, B. W.) 95–123 (Kluwer Academic, Dordrecht, 1994).
11. Rettallack, G. J. *Soils of the Past: Introduction to Paleopedology* (Unwin Hyman, Boston, 1990).
12. Naraoka, H., Ohtake, M., Maruyama, S. & Ohmoto, H. Non-biogenic graphite in 3.8-Ga metamorphic rocks from the Isua District, Greenland. *Chem. Geol.* **133**, 251–260 (1996).
13. Watanabe, Y., Naraoka, H., Wronkiewicz, D. J., Condie, K. C. & Ohmoto, H. Carbon, nitrogen, and sulfur geochemistry of Archean and Proterozoic shales from the Kaapvaal Craton, South Africa. *Geochim. Cosmochim. Acta* **61**, 3441–3459 (1997).
14. Des Marais, D. J. Isotopic evolution of the biogeochemical carbon cycle during the Proterozoic eon. *Org. Geochem.* **27**, 185–193 (1997).
15. Des Marais, D. J., Bauld, J., Palmisano, A. C., Summons, R. E. & Ward, D. M. in *Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 299–308 (Princeton Univ. Press, Lawrenceville, 1992).
16. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossil and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
17. Campbell, S. E. Soil stabilization by a prokaryotic desert crust: implications for Precambrian land biota. *Origin of Life and Evolution of the Biosphere* **9**, 335–347 (1976).
18. Nisbet, E. G. in *Early Precambrian Processes* (eds Coward, M. P. & Ries, A. C.) 27–51 (Special Publication, Geological Society of London, 1995).
19. Ohmoto, H. Evidence in pre-2.2 Ga paleosols for the early evolution of atmospheric oxygen and terrestrial biota. *Geology* **24**, 1135–1138 (1996).
20. Luque, F. J., Pasteris, J. D., Wopenka, B., Rodas, M. & Barrenechea, J. F. Natural fluid-deposited graphite: Mineralogical characteristics and mechanisms of formation. *Am. J. Sci.* **298**, 471–498 (1998).
21. Strauss, H. & Moore, T. B. in *Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 709–797 (Princeton Univ. Press, Lawrenceville, 1992).
22. Rittenberg, K. C., Goni, M. A. Phosphorus distribution, C:N:P ratios, and $\delta^{13}\text{C}_{\text{org}}$ in arctic, temperate, and tropical coastal sediments: tools for characterizing bulk sedimentary organic matter. *Mar. Geol.* **139**, 123–145 (1997).
23. Ohmoto, H. & Goldhaber, M. B. in *Geochemistry of Hydrothermal Ore Deposits* (ed. Barnes, H. L.) 517–612 (Wiley & Sons, New York, 1997).

Acknowledgements

We thank T. Nagase for TEM analyses, L. R. Kump, J. F. Kasting, H. Naraoka, K. Yamaguchi, S. Ono, Y. Huang, K. M. Towe and J. A. Lamberski for comments on an earlier manuscript, and R. Capo, B. Stewart, D. J. Des Marais, R. Buick and J. M. Hayes for discussions. This work was supported by grants to H.O. from the Japanese Ministry of Science and Education, NSF and NASA.

Correspondence and requests for materials should be addressed to Y.W. (e-mail: yumiko@essc.psu.edu).

Nutritional constraints in terrestrial and freshwater food webs

James J. Elser*, William F. Fagan*, Robert F. Denno†, Dean R. Dobberfuhl‡, Ayoola Folarin*, Andrea Huberty†, Sebastian Interlandi§, Susan S. Kilham§, Edward McCauley||, Kimberly L. Schulz¶, Evan H. Siemann# & Robert W. Sterner*

* Department of Biology, Arizona State University, Tempe, Arizona 85287, USA

† Department of Entomology, University of Maryland, College Park, Maryland, 20742, USA

§ School of Environmental Science, Engineering, and Policy, Drexel University, Philadelphia, Pennsylvania, 19104, USA

|| Ecology Division, Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

¶ Department of Environmental and Forest Biology, State University of New York, Syracuse, NY 13210, USA

Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77005, USA

* Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota 55108, USA

Biological and environmental contrasts between aquatic and terrestrial systems have hindered analyses of community and ecosystem structure across Earth's diverse habitats. Ecological stoichiometry^{1,2} provides an integrative approach for such analyses, as all organisms are composed of the same major elements (C, N, P) whose balance affects production, nutrient cycling, and food-web dynamics^{3,4}. Here we show both similarities and differences in the C:N:P ratios of primary producers (autotrophs) and invertebrate primary consumers (herbivores) across habitats. Terrestrial food webs are built on an extremely nutrient-poor autotroph base with C:P and C:N ratios higher than in lake particulate matter, although the N:P ratios are nearly identical. Terrestrial herbivores (insects) and their freshwater counterparts (zooplankton) are nutrient-rich and indistinguishable in C:N:P stoichiometry. In both lakes and terrestrial systems, herbivores should have low growth efficiencies (10–30%) when consuming autotrophs with typical carbon-to-nutrient ratios. These stoichiometric constraints on herbivore growth appear to be qualitatively similar and widespread in both environments.

The concept of a food web has been a central organizing theme in ecology ever since its classical development⁵. Ecologists are now actively seeking ways to integrate interspecific interactions in food webs with the functional processes of energy flow and material cycling^{6–9}. Here we use the perspective of ecological stoichiometry to analyse factors affecting energy and material flows at the autotroph–herbivore interface in terrestrial and freshwater ecosystems. Our analysis quantifies important divergences and convergences in the stoichiometric structure at the base of food webs in these diverse habitats.

The base of terrestrial and freshwater food webs differed dramatically in C:nutrient ratios (Fig. 1). Mean C:N and C:P ratios of the foliage of terrestrial autotrophs were more than threefold higher than for freshwater seston (lake particulate matter, generally dominated by phytoplankton) (*t*-tests, *P* < 0.0001), indicative of the influence of nutrient-poor, C-rich structural carbohydrates in vascular plant tissue. Thus, a generalist terrestrial herbivore consuming a plant with the average plant C:nutrient ratio would acquire less than a third as many nutrient atoms per C atom ingested as would a freshwater zooplankton feeding on average seston. Furthermore, C:N ratios varied more in the terrestrial data: the coefficient of variation (c.v.) of terrestrial autotroph biomass was 0.64 but only 0.29 for freshwater seston. Whereas autotroph C:nutrient ratios diverged widely between terrestrial and freshwater ecosystems, autotroph N:P ratios did not (Fig. 1; 28 for terrestrial foliage versus 30 for freshwater seston; *t*-test, *P* = 0.65). Furthermore, the biomass N:P ratio did not vary as greatly between terrestrial and freshwater habitats as did the C:N ratio (c.v. of N:P was 0.54 versus 0.53 for C:N). Thus, despite major differences in the size, complexity and taxonomic affiliation of autotrophic organisms in terrestrial and freshwater realms, patterns of biomass composition for N and P were very similar. Given the prevailing view that in general primary production is limited by P in freshwater ecosystems but by N in terrestrial systems¹⁰, the coincidence of autotroph biomass N:P in terrestrial and freshwater systems is intriguing (Fig. 1). This similarity may indicate that the prevalence of N-limitation in lakes is greater than previously thought¹¹ or that P-limitation in terrestrial systems is more widespread than generally acknowledged. Or, it is possible that biomass N:P values indicative of N versus P limitation differ for freshwater and terrestrial autotrophs, but this is not supported by recent investigations in wetland vegetation¹².

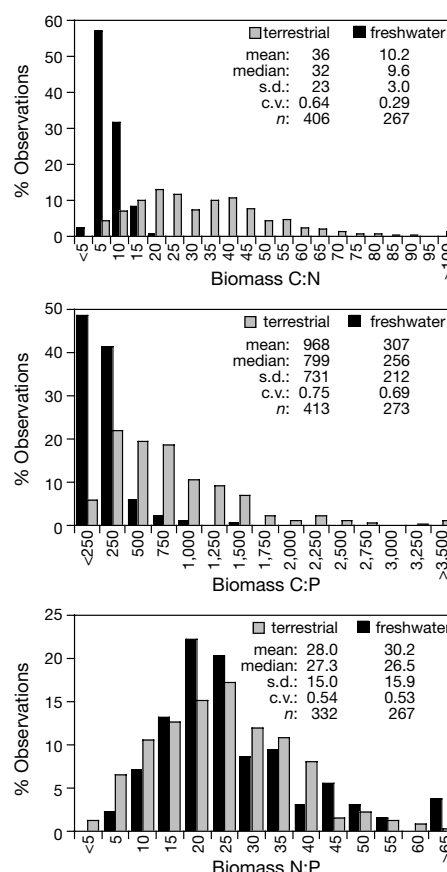


Figure 1 Frequency histograms summarizing C:N:P stoichiometry in autotrophs at the base of terrestrial and freshwater food webs. All stoichiometric ratios are atomic.

† Present address: St Johns River Water Management District, Palatka, Florida 32177, USA.

Like autotrophs, metazoan herbivores in terrestrial and freshwater habitats differ considerably in size and taxonomic affiliation. Despite these contrasts, we found that herbivorous insects and zooplankton had similar mean C:N, C:P and N:P ratios in the two habitat types (Fig. 2; t -tests, $P > 0.17$). In both groups C:P ratios varied more than C:N ratios among taxa, consistent with the relative constancy of N content but wide variation of P content in major biological molecules (that is, N-rich proteins versus N- and P-rich nucleic acids) and structures². Associations of body C:N:P with body growth rate due to the necessity of increased allocation of resources to P-rich ribosomal RNA² have been documented in planktonic crustaceans^{2,13}. We show that as much variation in body C:N:P ratios exists among insect herbivores as among zooplankton. However, connections between patterns of elemental and biochemical composition and growth rate in insects remain largely unexplored.

The C:N and C:P ratios of the herbivores were considerably lower than those of their potential foods, especially in terrestrial systems where C:nutrient ratios of foliage exceeded insect C:nutrient ratios by more than 5- to 10-fold on average. This indicates that stoichiometric food quality for herbivores is generally poor in lakes and especially in terrestrial habitats. The extreme imbalance in C:nutrient ratios between plant and insect biomass may contribute to the prevalence of feeding specialization among insect herbivores^{14,15}, in which insects focus their consumption on those plant species, tissues, or time-frames for which food nutrient content is more suitable. In contrast, pelagic herbivores in lakes appear to face less daunting constraints and may thus succeed with relatively indiscriminate, filter-feeding modes of nutrition. Given elevated C:nutrient ratios, which element (N or P) is likely to be in shorter supply relative to herbivore needs? Mean body N:P of herbivore taxa (~24) was lower than that of average autotroph biomass (~30;

t -test, $P = 0.02$), indicating that, on average, N is in excess relative to P for herbivores in these systems. Yet, because herbivore and autotroph N:P vary among species, the identity of the potentially limiting constituent will depend on the relative N:P of particular resource-herbivore combinations. However, the fact that plant N:P was significantly higher than insect N:P suggests that greater attention should be paid to the potential for inadequate P intake for herbivores in terrestrial ecosystems, where ecologists have focused primarily on N¹⁶.

Our data clearly show that the stoichiometry of the autotroph-herbivore interaction is greatly unbalanced in freshwater and especially terrestrial ecosystems. For example, in freshwater zooplankton, *Daphnia* are relatively P-rich with body C:P ratios of about 80:1⁴. According to stoichiometric food quality models and experimental investigations^{3,17,18}, *Daphnia* face potential P-deficiency when food C:P exceeds ~250, a situation that appears commonplace in lakes (Fig. 1). Our data indicate that terrestrial herbivores with a P-rich lifestyle similar to *Daphnia*'s encounter even more daunting constraints: only 6% of the autotroph species sampled in our data set had plant C:P less than 250:1. Stoichiometric theory predicts that homeostatic herbivores consuming elementally imbalanced food will exhibit strongly diminished efficiency of conversion of ingested carbon into new biomass (for example, reduced "gross growth efficiency", GGE_C). Indeed, examples of terrestrial and freshwater herbivores (caterpillars of *Pieris rapae*¹⁹ and *Daphnia magna*²⁰, respectively) show the reduction in GGE_C with food C:nutrient ratio (Fig. 3) predicted by stoichiometric food quality models^{17,18}. Comparison of these responses with the histograms in Fig. 1 indicates that such herbivores would exhibit low GGE_C when consuming average autotroph biomass in their respective habitats (30% for *Daphnia*, <10% for *Pieris*). Given the possible connection between animal C:N:P ratios and growth rate², such food quality limitations may fall disproportionately on fast-growing 'outbreak' herbivores that require nutrient-rich resources

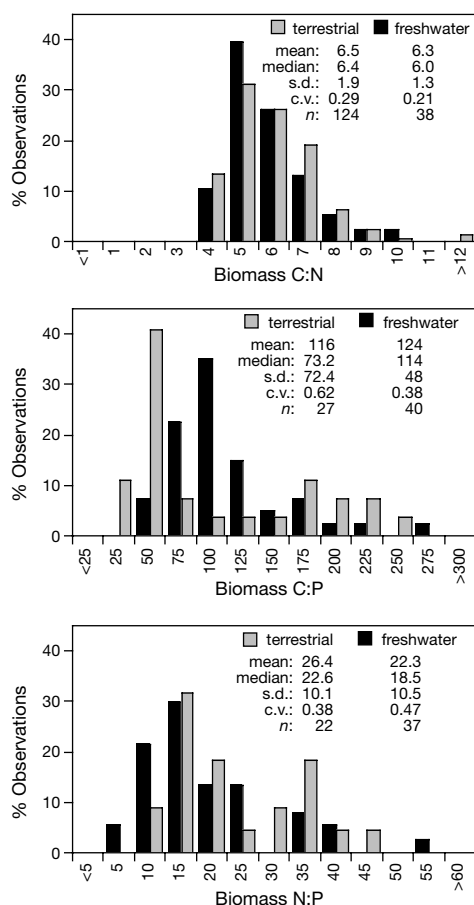


Figure 2 Frequency histograms summarizing C:N:P stoichiometry of invertebrate herbivores in terrestrial and freshwater habitats. All stoichiometric ratios are atomic.

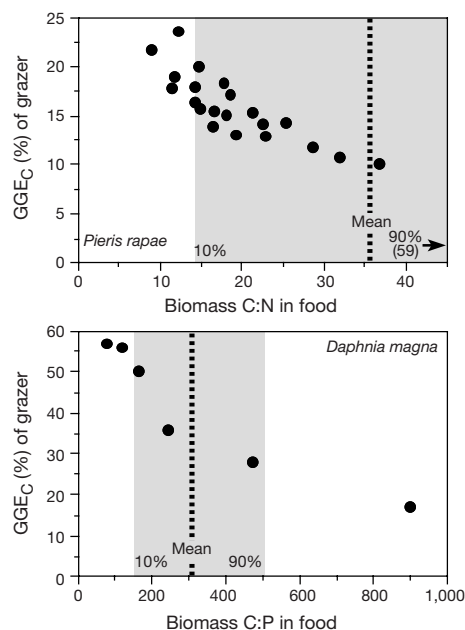


Figure 3 Decline in gross growth efficiency (GGE_C) for typical terrestrial and freshwater grazers with increasing food carbon-to-nutrient ratio. The grazers were cabbage butterfly caterpillars, *Pieris rapae*¹⁹ (terrestrial) and the zooplankter *Daphnia magna*²⁰ (freshwater). Data on GGE_C (proportion of ingested carbon successfully converted to body growth) and food nutrient content were extracted from the cited studies. Percentage N in the caterpillar study was converted to C:N assuming a percentage C value of 46.4%. Responses are superimposed on observed values of C:N in foliage and C:P in seston; for each, the dashed line indicates the mean value and shading indicates the 10% and 90% limits from the frequency distribution (Fig. 1).

to rapidly build nutrient-rich bodies. In general, our analyses show large differences in C:nutrient balance across the autotroph–herbivore interface in freshwater and terrestrial ecosystem that may have important effects on the intensity of herbivory^{21,22} and the fate of organic matter²³ in diverse ecosystems. However, terrestrial and aquatic food webs share great similarity in the N:P stoichiometry of autotroph–grazer interactions. □

Methods

Autotrophs

We supplemented the literature with unpublished data to develop databases documenting the C:N:P stoichiometry of terrestrial plants and suspended particulate matter ('seston') in lakes. We restricted our terrestrial autotroph database to elemental analyses of foliage collected under field conditions, excluding agricultural and greenhouse studies. Multiple data for a single species were averaged before analysis. Terrestrial data were most frequently reported in percentage dry weight terms (% N, % P); when % C values were not reported we converted % N and % P data to C:N and C:P ratios using the mean percentage C of reported values (46.4% C). To evaluate whether this procedure introduced any bias to observed patterns in foliage C:P and C:N, we calculated the mean and variability (coefficient of variation, c.v.) of the C:N and C:P ratios for that subset of species for which % C, % N, and % P were all reported ($n = 44$). The mean and c.v. of C:N for this limited data set were 35.9 and 0.57, respectively, and 805 and 0.78 for C:P. These values are reasonably close to those for the remaining entries (for C:N, mean was 36.5, c.v. was 0.64; for C:P, mean was 990, c.v. was 0.75); thus, using a fixed percentage C value to estimate C:N and C:P probably did not influence the major patterns observed. A total of 501 plant species from 358 genera, 107 families, 62 orders, 20 subclasses, 8 classes and 5 divisions were included. We assessed C:N:P stoichiometry at the base of freshwater pelagic food webs by compiling a database of seston elemental composition in 226 lakes from published and unpublished reports. Only data for surface waters during the summer growing season were included; multiple observations during a year were averaged, and thus a 'lake-year' was the primary observation unit. Data were generally for lakes of small to moderate size but information for several of the world's great lakes was also included. Lakes were primarily located in North America but seston data for lakes in Europe, Africa and Asia were also obtained. Seston contains a mixture of living algae but also bacteria, protozoa and detritus and forms the food base for relatively indiscriminate planktonic filter-feeders. Although the contribution of these different components probably differs among lakes, various data indicate that, in general, seston particles in stratified lakes are dominated by phytoplankton biomass. For example, even in some lakes where seston C:P was high (and thus the contribution of low-nutrient detritus might be thought relatively important), algae contributed about 70% of total seston biomass (bacteria and protozoa contributed ~20% and <5%, respectively, implying little influence of detritus)²⁴. Thus, the freshwater and terrestrial data sets for 'autotrophs' differ in that the terrestrial data involve observations for particular plant species while the lake data correspond to a mixture of particles, living and non-living. Finally, if different seston particles have substantially different C:N:P ratios, bulk seston C:N:P measurements may not accurately quantify actual stoichiometric food quality for particular herbivores that can discriminate among particles, such as some calanoid copepods⁵.

Herbivores

Data for the C:N:P stoichiometry of terrestrial herbivorous insects and lake zooplankton were compiled from published and unpublished sources. Multiple data for a single species were averaged before analysis. As for terrestrial plants, when values of percentage C were not given, data reported as % N and % P were converted to C:N and C:P ratios using the mean percentage C value for the remainder of the herbivore database (48% C). We followed the same procedure used in analyses of the foliage data to evaluate possible bias introduced by assuming this fixed percentage C value. However, data for few species included all three parameters (% C, % N, % P); we thus confined our assessment of possible biases to data on herbivore C:N. The mean and c.v. values of C:N for the data subset with direct measurements of % C and % N ($n = 67$) were 5.9 and 0.21 whereas values for entries for which the fixed percentage C value was used ($n = 97$) were 6.7 and 0.28. Here again, using a fixed value of percentage C to estimate C:N and C:P from % N and % P probably did not unduly influence the observed patterns. A total of 130 species of insects from 93 genera, 40 families and 7 orders were included. By far, most insects included were leaf-eating, though a minority were phloem-feeding herbivores (such as aphids). Leaf-eaters and phloem-feeders did not differ in C:N:P ratios and therefore all taxa were analysed together. Predatory zooplankton were excluded from the compilation but several omnivorous taxa were retained. A total of 43 species of zooplankton from 23 genera, 12 families, 8 orders, 4 classes and 2 phyla were included. The majority of the taxa were crustaceans (mainly branchiopods ('cladocera'), malacostracans and copepods) but data for several rotifers were also compiled. All stoichiometric ratios were calculated on an atomic basis. A complete summary of the data sets, including original citations, can be obtained at <http://www.nceas.ucsb.edu/ecostochiometry>.

Received 31 March; accepted 1 September 2000.

- Reiners, W. A. Complementary models for ecosystems. *Am. Nat.* **127**, 59–73 (1986).
- Elser, J. J., Dobberfuhl, D., MacKay, N. A. & Schampel, J. H. Organism size, life history, and N:P stoichiometry: towards a unified view of cellular and ecosystem processes. *BioScience* **46**, 674–684 (1996).
- Sterner, R. W. & Hessen, D. O. Algal nutrient limitation and the nutrition of aquatic herbivores. *Annu. Rev. Ecol. Syst.* **25**, 1–29 (1994).

- Elser, J. J. & Urabe, J. The stoichiometry of consumer-driven nutrient cycling: theory, observations, and consequences. *Ecology* **80**, 735–751 (1999).
- Elton, C. *Animal Ecology* (Sidgwick and Jackson, London, 1927).
- Chapin, F. S. *et al.* Biotic control over the functioning of ecosystems. *Science* **277**, 500–504 (1997).
- Leibold, M. A. Resource edibility and the effects of predators and productivity on the outcome of trophic interactions. *Am. Nat.* **6**, 922–949 (1989).
- Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).
- McCauley, E., Nisbet, R. M., Murdoch, W. W., DeRoos, A. M. & Gurney, W. S. C. Large amplitude cycles of *Daphnia* and its algal prey in enriched environments. *Nature* **402**, 653–656 (1999).
- Schlesinger, W. H. *Biogeochemistry: An Analysis of Global Change* (Academic, San Diego, 1997).
- Elser, J. J., Marzolf, E. R. & Goldman, C. R. Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: a review and critique of experimental enrichments. *Can. J. Fish. Aquat. Sci.* **47**, 1468–1477 (1990).
- Verhoeven, J. T. A., Koerselman, W. & Meuleman, A. F. M. Nitrogen- or phosphorus-limited growth in herbaceous wet vegetation: relations with atmospheric inputs and management regimes. *Trends Ecol. Evol.* **11**, 494–497 (1996).
- Main, T., Dobberfuhl, D. R. & Elser, J. J. N:P stoichiometry and ontogeny in crustacean zooplankton: a test of the growth rate hypothesis. *Limnol. Oceanogr.* **42**, 1474–1478 (1997).
- Futuyma, D. J. & Gould, F. Associations of plants and insects in a deciduous forest. *Ecol. Monogr.* **49**, 33–50 (1979).
- Strong, D. R., Lawton, J. H. & Southwood, R. *Insects on Plants: Community Patterns and Mechanisms* (Blackwell Scientific, London, 1984).
- White, T. C. R. *The Inadequate Environment: Nitrogen and the Abundance of Animals* (Springer, New York, 1993).
- Andersen, T. *Pelagic Nutrient Cycles: Herbivores as Sources and Sinks* (Springer, Berlin, Heidelberg & New York, 1997).
- Sterner, R. W. Modelling interactions between food quality and quantity in homeostatic consumers. *Freshwat. Biol.* **38**, 473–482 (1997).
- Slansky, F. & Feeny, P. Stabilization of the rate of nitrogen accumulation by larvae of the cabbage butterfly on wild and cultivated plants. *Ecol. Monogr.* **47**, 209–228 (1977).
- DeMott, W. R., Gulati, R. D. & Siewertsen, K. Effects of phosphorus-deficient diets on the carbon and phosphorus balance of *Daphnia magna*. *Limnol. Oceanogr.* **43**, 1147–1161 (1998).
- Cyr, H. & Pace, M. L. Magnitude and patterns of herbivory in aquatic and terrestrial ecosystems. *Nature* **361**, 148–150 (1993).
- Strong, D. R. Are trophic cascades all wet? Differentiation and donor-control in speciose ecosystems. *Ecology* **73**, 747–754 (1992).
- Cebrian, J. Patterns in the fate of production in plant communities. *Am. Nat.* **154**, 449–468 (1999).
- Elser, J. J., Chrzanowski, T. H., Sterner, R. W., Schampel, J. H. & Foster, D. K. Elemental ratios and the uptake and release of nutrients by phytoplankton and bacteria in three lakes of the Canadian Shield. *Microb. Ecol.* **29**, 145–162 (1995).
- Butler, N. M., Suttle, C. A. & Neill, W. E. Discrimination by freshwater zooplankton between single algal cells differing in nutritional status. *Oecologia* **78**, 368–372 (1989).

Acknowledgements

This paper is a contribution from the Ecological Stoichiometry working group at the National Center for Ecological Analysis and Synthesis, a centre funded by the National Science Foundation, the University of California, and the State of California. We thank the staff of NCEAS for logistical support. We also thank S. Nielsen for providing an electronic summary of his extensive data set on autotroph elemental composition. D. Strong, I. Loladze and C. Mitter provided useful comments.

Correspondence and requests for materials should be addressed to J.E. (e-mail: j.elsner@asu.edu).

Bacterial dehalorespiration with chlorinated benzenes

Lorenz Adrian*†, Ulrich Szewzyk†, Jörg Wecke‡ & Helmut Görisch*

* Fachgebiet Technische Biochemie and † Fachgebiet Ökologie der Mikroorganismen, Technische Universität, 13353 Berlin, Germany
‡ Robert Koch Institut, 13353 Berlin, Germany

Chlorobenzenes are toxic, highly persistent and ubiquitously distributed environmental contaminants that accumulate in the food chain¹. The only known microbial transformation of 1,2,3,5-tetrachlorobenzene (TeCB) and higher chlorinated benzenes is the reductive dechlorination to lower chlorinated benzenes under anaerobic conditions observed with mixed bacterial cultures^{2–4}. The lower chlorinated benzenes can subsequently be mineralized by aerobic bacteria. Here we describe the isolation of the oxygen-sensitive strain CBDB1, a pure culture capable of reductive

dechlorination of chlorobenzenes. Strain CBDB1 is a highly specialized bacterium that stoichiometrically dechlorinates 1,2,3-trichlorobenzene (TCB), 1,2,4-TCB, 1,2,3,4-TeCB, 1,2,3,5-TeCB and 1,2,4,5-TeCB to dichlorobenzenes or 1,3,5-TCB. The presence of chlorobenzene as an electron acceptor and hydrogen as an electron donor is essential for growth, and indicates that strain CBDB1 meets its energy needs by a dehalorespiratory process. According to their 16S rRNA gene sequences, strain CBDB1, *Dehalococcoides ethenogenes*⁵ and several uncultivated bacteria form a new bacterial cluster, of which strain CBDB1 is the first, so far, to thrive on a purely synthetic medium.

The reductive dehalogenation of chlorinated aromatic or aliphatic substances has been identified as an energy-yielding process in a number of anaerobic bacteria⁶. The reductively dechlorinating bacteria known to date belong to the low GC Gram-positive bacteria (*Desulfotobacterium* and *Dehalobacter*) or to the *Proteobacteria* (for example, *Desulfomonile*, *Desulfuromonas* or *Dehalospirillum*), whereas *Dehalococcoides* was described as having no close affiliation^{5,6} or as being affiliated to the green non-sulphur bacteria (refs 7, 8; see also <http://www.ncbi.nlm.nih.gov>). Pure strains have been isolated that reductively dechlorinate chlorobenzoates, chlorophenols, bromophenols or chlorinated ethenes; however, bacteria that have an energy metabolism relying on reductive dechlorination of chlorobenzenes have not been described.

For all multiple-chlorinated benzenes, stoichiometric reductive dechlorination to lower-chlorinated benzenes has been reported in mixed cultures^{3,9}. Different routes of chlorobenzene dechlorination and other physiological characteristics of enriched mixed cultures have been determined^{10–13}. We describe the first isolation, to our knowledge, of a bacterium thriving on the energy obtained from reductive dehalogenation of chlorobenzenes.

Isolation was started from a previously characterized¹² and enriched chlorobenzene-dechlorinating mixed culture¹³. Growth of the dechlorinating bacteria in the mixed culture strictly depends on the concentration of TCB added, indicating that energy conservation is linked to reductive dechlorination. Isolation of dechlorinating bacteria from the consortium could not be achieved by random isolation of predominant bacteria¹³. We therefore established a cultivation system that allows the determination of dechlorinating activity in solidified media. Dichlorobenzene (DCB) formation by the enriched consortium was found in media containing 0.3% of a low-melting agarose but not in media with standard agar or agarose. 1,2,3- and 1,2,4-TCB were reduced completely to a mixture of 1,3- and 1,4-DCB after 2–3 weeks of incubation in solid media. The first transfers were performed in a medium containing 30 μ M TCB, 5 mM butyrate, 5 mM succinate, 5 mM formate, 1.6 mM citrate, 1 mM sulphate and 20% (v/v) of a sterile supernatant from mixed cultures in liquid medium. Under these conditions up to 99.3% of the TCB added was converted to DCB within 20 days. However, visible colony growth was not detected. By using chemical analysis, we identified the main organic components in the sterile supernatant as formate and acetate, which were formed from citrate by Gram-positive bacteria in the mixed consortium. Dechlorinating activity and colony growth were observed in Ti(III)-citrate-reduced agarose shake cultures of the mixed consortium with 30 μ M TCB, 5 mM acetate, 5 mM formate and 30–80 μ M glucose. Addition of hydrogen enhanced the dechlorinating activity and allowed us to omit glucose and formate from the medium; also, Ti(III) citrate could be replaced by Ti(III) nitrilotriacetate (NTA).

Cultures restricted to TCB, hydrogen and acetate showed high dechlorinating activity and also faint colony growth after 3 weeks of incubation. The colonies were subcultured six successive times in solidified medium with TCB, hydrogen and acetate, maintaining high dechlorinating activity. After these isolation steps, no contaminating bacteria could be detected by cultivation in anaerobic agarose shake cultures with 5 mM pyruvate, 5 mM glucose or 50 mg l⁻¹

yeast extract, or by aerobic cultivation on nutrient broth agar plates. The isolate was designated strain CBDB1, and is very oxygen sensitive. Exposure to aerobic conditions for a few seconds results in a complete and irreversible loss of dechlorinating activity.

Growth of strain CBDB1 to a faint colony size was achieved in Ti(III)-NTA- or Ti(III)-citrate-reduced medium containing 30 μ M TCB, 3.75 mM hydrogen (nominal concentration) and 5 mM acetate. No growth could be detected when TCB, hydrogen or acetate was omitted from the medium. With TCB and hydrogen, but omitting acetate, dechlorination to DCB occurred, but successive transfers resulted in loss of the dechlorinating activity. Formate could not replace hydrogen as an electron donor for TCB dechlorination by strain CBDB1.

There was no fermentative growth of strain CBDB1 in solidified media with 50, 100, 500 or 1,000 μ M pyruvate, 50 μ M glucose, 50 μ M sucrose, 50 μ M fructose, 50 μ M galactose, 1 mM citrate or 5 mM succinate. High-performance liquid chromatography showed

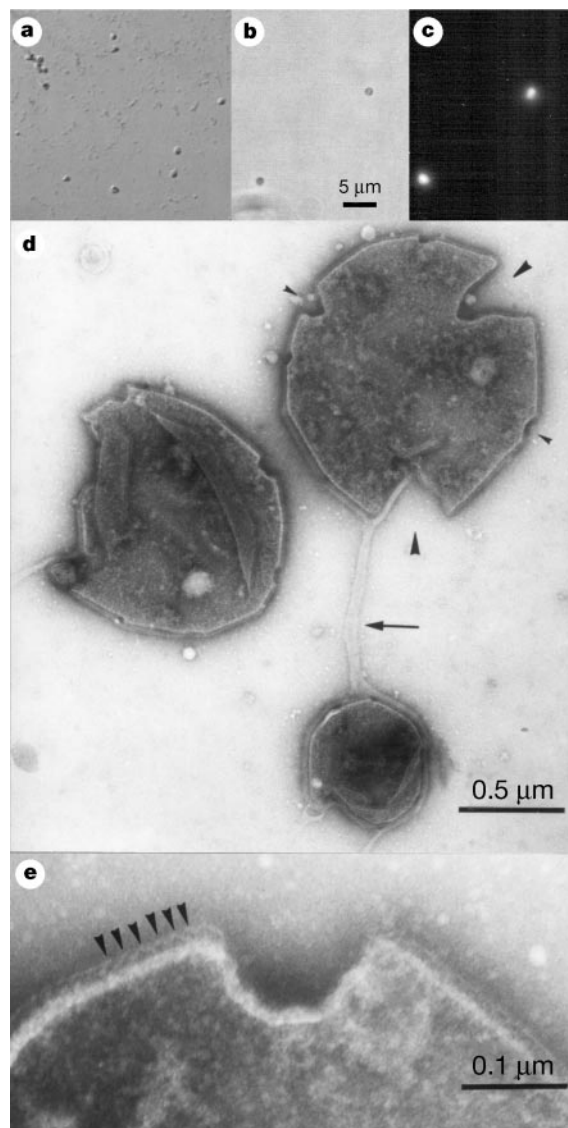


Figure 1 Micrographs of strain CBDB1. **a**, Differential interference contrast after staining with 1% safranin. **b**, Phase contrast. **c**, Epifluorescence after staining with DAPI (1 μ g ml⁻¹); same object as in **b**. **d**, Electron micrograph of CBDB1 cells after negative staining with 1% phosphotungstic acid. Bacteria lack a typical cell wall but show notches on the surface (big and small arrow heads) and an appendix (arrow) enclosing another vesicular structure. **e**, Higher magnification of the peripheral part of a CBDB1 cell including a notch after negative staining. The outer sheath with regularly arranged subunits is shown (arrowheads).

that citrate added as a chelating agent for Ti(III) was not transformed. We did not detect methanogenic activity with acetate nor with hydrogen plus carbonate, nor did we detect acetogenesis from hydrogen and carbonate.

Growth to a visible colony size did not occur in solidified medium with 1 mM concentrations of sulphate, sulphite, thiosulphate, nitrate, fumarate or 3-chloro-4-hydroxy benzoate as an electron acceptor using hydrogen (3.75 mM nominal concentration) as an electron donor and acetate (5 mM) as a carbon source. Addition of 1 mM thiosulphate, sulphite or nitrate to medium with TCB, hydrogen and acetate led to a complete inhibition of DCB formation, whereas sulphate, fumarate or 3-Cl-4-hydroxybenzoic acid did not alter the rate of DCB formation. Sulphide production was not detected from sulphate, sulphite or from thiosulphate.

Growth of strain CBDB1 was therefore only possible when chlorobenzenes, hydrogen and acetate were supplied in the medium. Growth in a completely synthetic medium with TCB, hydrogen and acetate, however, shows that energy conservation occurred by means of a respiratory process, because these substrates

cannot be used for substrate-level phosphorylation.

In a completely defined liquid medium containing 30 μ M TCB, 5 mM acetate and 7.5 mM hydrogen (nominal concentration), strain CBDB1 could be subcultured indefinitely using either Ti(III) citrate or Ti(III) NTA as a reducing agent. Vitamin B₁₂ was necessary to maintain high dechlorinating activities: no DCB was formed after 12 days of incubation without vitamin B₁₂ but control cultures with vitamin B₁₂ dechlorinated 60% of the supplied TCB. No DCB was formed in controls in the presence of vitamin B₁₂ and autoclaved inoculum. As observed with the mixed culture¹², the pure culture first dechlorinated 15 μ M 1,2,3-TCB to 1,3-DCB. 1,2,4-TCB was dechlorinated to a mixture of 72% 1,4-DCB and 28% 1,3-DCB ($\pm$ 2%) only after the concentration of 1,2,3-TCB dropped to below the detection limit of 0.5 μ M. However, complete dechlorination of TCB to DCB required 14–21 days, which is longer than the 10–12 days that were needed for complete dechlorination of TCB to DCB by the mixed culture. This suggests that other bacteria in the mixed culture contributed to favourable conditions for strain CBDB1. Bacterial cell counts after complete dechlorination were about 2×10^6 cells per ml as estimated by direct cell counting. Cultures with 5 mM acetate, 7.5 mM hydrogen (nominal concentration) and 30 μ M TCB were amended with a further 10 mM 1,2,3-TCB (nominal concentration) in a hexadecane phase¹⁰ after 3 weeks of incubation when all of the initially added TCB was dechlorinated. In these cultures, 5–7 mM TCB were dechlorinated within an additional 2 weeks of incubation, resulting in cell numbers of about 10^7 cells per ml.

A seven-week incubation period in a two-liquid-phase hexadecane/water system was used to determine the dechlorinating activity of strain CBDB1 towards other chlorinated benzenes. We did not observe dechlorination of DCB or 1,3,5-TCB. 1,2,3,4-TeCB and 1,2,4,5-TeCB were both transiently converted to 1,2,4-TCB, resulting in the formation of 1,3- and 1,4-DCB. 1,2,3,5-TeCB was exclusively dechlorinated to 1,3,5-TCB. Dechlorination of pentachlorobenzene to a mixture of 1,3-DCB, 1,4-DCB and 1,3,5-TCB with the transient formation of 1,2,4,5-TeCB, 1,2,3,5-TeCB and 1,2,4-TCB was detected with the highly enriched consortium from which strain CBDB1 was isolated; however, dechlorination of pentachlorobenzene or hexachlorobenzene could not be shown with the pure strain CBDB1.

We performed several series of experiments in homogeneous medium containing 5 mM acetate, 7.5 mM hydrogen (nominal concentration) and 0.55 mM perchloroethene. A portion of (10–20%) of the PCE that we added was abiotically dechlorinated to trichloroethene in one week by Ti(III). Microbial dechlorination of perchloroethene or the abiotically formed trichloroethene by strain CBDB1 could not be shown under the cultivation conditions that we used.

Strain CBDB1 formed DCB from TCB in the presence of up to 0.5 g l⁻¹ vancomycin or up to 10 g l⁻¹ ampicillin. A culture with high dechlorinating activity was maintained for five successive transfers in the presence of 0.5 mg l⁻¹ vancomycin, showing that strain CBDB1 was able to grow in the presence of the antibiotic.

Differential interference and phase-contrast microscopy of samples from a culture grown with 1,2,3-TCB (10 mM nominal concentration added as a 0.2 M solution in hexadecane), hydrogen (10 mM nominal concentration) and acetate (5 mM) revealed irregular coccoid bacteria with a diameter of 1–2 μ m (Fig. 1a–c). On electron micrographs, cells without a typical bacterial cell wall are visible (Fig. 1d, e), and there are often appendices that may enclose other vesicular structures (Fig. 1d). Moreover, the negatively stained cells show notches of different size. Notches positioned opposite each other show roughly the same size. Such notches may reflect consecutive division planes of a coccus¹⁴. Instead of a typical bacterial cell wall, an S-layer-like structure showing a regular periodicity is covering the surface of the cells (Fig. 1e).

A stretch of 1,421 base pairs of the 16S rRNA gene sequence of

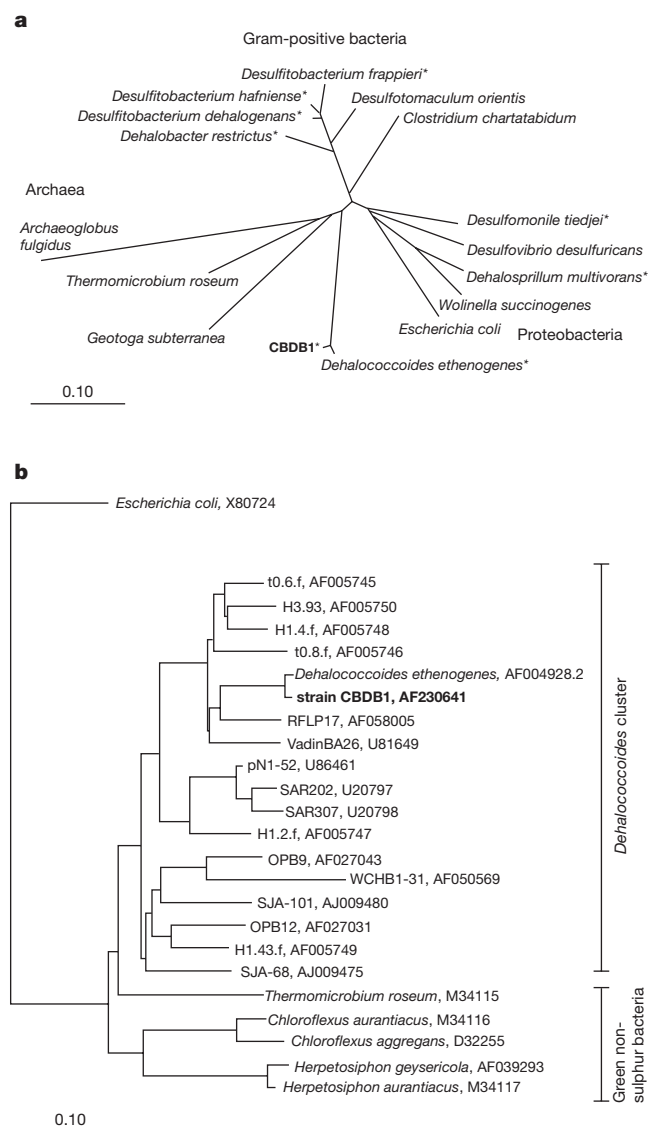


Figure 2 Phylogenetic affiliation of strain CBDB1 based on currently available 16S rRNA gene sequences. **a**, General prokaryotic tree. Other reductively dechlorinating organisms (indicated by an asterisk) and reference organisms are shown. **b**, Tree of sequences from the *Dehalococcoides* cluster and from the green non-sulphur bacteria. Sequences from uncultivated bacteria are labelled with the sequence designation and GenBank accession numbers are supplied.

strain CBDB1 was determined. The sequence shares 1,398 identical base pairs (98.3% of the alignment) with the 16S rRNA gene of *D. ethenogenes* (GenBank accession no. AF004928.2)⁵. All other reported sequences have less than 90.8% sequence similarity. On the basis of available data, strain CBDB1 and *D. ethenogenes* form a new branch phylogenetically distant from all other dechlorinating bacteria (Fig. 2a). According to calculations with the ARB software¹⁵, this new branch is loosely affiliated with a larger group of sequences of uncultivated, and therefore unknown, bacteria detected by cloning approaches (Fig. 2b). This cluster includes sequences determined from a 16S ribosomal DNA clone library⁸ obtained from the same anaerobic TCB-dechlorinating reactor that we used here as the initial inoculum to isolate strain CBDB1. The cluster also contains 16S rDNA sequences obtained from other dehalogenating mixed cultures and sequences from contaminated natural sites. The findings suggest that the bacteria of the *Dehalococcoides* cluster provide important functions in contaminated environments⁸. In this newly described *Dehalococcoides* cluster, strain CBDB1 is the first bacterium that grows on a purely synthetic medium, which allows its comprehensive physiological and biochemical characterization.

The separated phylogenetic position of strain CBDB1 and *D. ethenogenes* is reflected by their unique physiological properties. In both organisms, reductive dehalogenation of chlorinated compounds is the only energy-conserving process found to date. Neither of the organisms grows by fermentation nor by respiration using non-chlorinated terminal electron acceptors. Furthermore, as *D. ethenogenes* is the only known isolate to dechlorinate perchloroethene to the non-toxic end-product ethene⁵, strain CBDB1 is the only known isolate that uses reductive dechlorination of chlorobenzenes for energy conservation. Both organisms seem to be restricted to the use of hydrogen as an electron donor. The resistance of *D. ethenogenes* to vancomycin up to 0.1 g l⁻¹ and to ampicillin up to 3 g l⁻¹, hybridization experiments with fluorescently labelled wheat germ agglutinin, and electron microscopy⁵ all indicate that *D. ethenogenes* has an unusual cell-wall structure and composition. Resistance against high concentrations of vancomycin and ampicillin was also found for strain CBDB1. This resistance, coupled with the electron microscopy, suggests that strain CBDB1 has no typical peptidoglycan cell wall, but has another type of cell envelope with characteristic notches.

Although the cultivation of *D. ethenogenes* requires the addition of undefined supplements⁵, cultivation of strain CBDB1 was achieved in an entirely synthetic medium. The methods developed may aid in the cultivation, isolation and characterization of related bacteria with further dehalogenating potential towards highly chlorinated aromatic compounds. □

Methods

Media, cultivation and isolation

We used a completely defined mineral medium containing vitamins, 15 µM 1,2,3-TCB and 15 µM 1,2,4-TCB¹². Isolation of strain CBDB1 was achieved by dilution in solidified medium under anaerobic conditions¹⁶. Glass tubes with a volume of 15 ml were filled with 2 ml of 20 mg ml⁻¹ Seaplaque agarose low melting (FMC) and autoclaved. We added 10 ml medium (32 °C) and substrates to the molten agarose. Where indicated, we replaced 2 ml of basal medium with 2 ml of sterile supernatant that we obtained from mixed dechlorinating cultures in liquid medium¹³. We added 120 µl of Ti(III) citrate solution¹⁷ (0.1 M Ti(III), 0.2 M citrate) or Ti(III) NTA¹³. Tubes were closed with butyl rubber septa and screw caps. A gas phase of N₂/CO₂ (80/20%, v/v) was supplied through 0.45-µm injection needles using a specially developed and self-constructed gas exchange device controlled by a microprocessor. The device allows a reliable and reproducible gas exchange in several tubes. Where indicated, we injected hydrogen with a gas-tight syringe. Tubes were shaken and incubated at 32 °C for 30 min to establish stable redox conditions, before inoculation and dilution with sterile glass syringes. After incubation between one and several weeks at 29 °C in the dark without shaking, colonies were picked with sterile Pasteur pipettes partly filled with suspension buffer (130 mM NaCl, 10 mM Na₂HPO₄ / NaH₂PO₄, pH 7.2, 0.5 mg l⁻¹ resazurin and 1 mM Ti(III)). We transferred colonies to 2-ml glass flasks containing 2 ml of suspension buffer and a stirring bar. After closing the flasks with butyl rubber septa and screw caps, the bacterial suspensions were stirred for 10 min and afterwards used for inoculation of fresh dilution series or liquid medium.

Strain CBDB1 was cultivated in 60-ml flasks containing 30 ml of liquid medium with 5 mM acetate and 240 µl of Ti(III) citrate or Ti(III) NTA solution and 5 ml of hydrogen in the gas phase, using an 0.5% inoculum and the procedures described for tubes. Where indicated, chlorobenzenes were added as a 0.2 M solution in hexadecane after inoculation. Tests for aerobic growth were done on agar plates containing 12 g l⁻¹ agar (Oxoid) and 8 g l⁻¹ nutrient broth (Difco).

Light and electron microscopy

We used an Axioskop (Zeiss) for light microscopy and photomicrographs. We took photomicrographs from cell suspensions previously treated by centrifugation (10,000g, 10 min); we took electron micrographs of cell suspensions after negative staining with 1% phosphotungstic acid at pH 6.8. We visualized the negative-stained cells with a Phillips 400 electron microscope.

Analysis

Chlorobenzenes were extracted from liquid or solidified medium with hexane¹² and quantified by gas chromatography (Shimadzu 14B) using flame ionization detection and a Permabond-FFAP capillary column (Macherey & Nagel). We analysed chlorinated ethenes by gas chromatography and electron capture detection. We determined formate, acetate and citrate concentrations by high-performance liquid chromatography¹³, we quantified sulphide by gas chromatography with thermal conductivity detection¹².

Determination of the 16S rRNA gene sequence

The 16S rRNA gene sequence was determined from 10 ml of a pure culture by polymerase chain reaction¹⁸ at 55 °C primer hybridization temperature, and by subsequent cycle sequencing¹⁹. We performed phylogenetic analysis with the ARB software package¹⁵ as described¹⁸. We checked results with the BLAST tool at NCBI (<http://www.ncbi.nlm.nih.gov/blast/>).

Received 9 June; accepted 6 September 2000.

1. Oliver, B. G. & Nicol, K. D. Chlorobenzenes in sediments, water, and selected fish from lakes Superior, Huron, Erie, and Ontario. *Environ. Sci. Technol.* **16**, 532–536 (1982).
2. Tiedje, J. M., Boyd, S. A. & Fathepure, B. Z. Anaerobic degradation of chlorinated aromatic hydrocarbons. *J. Ind. Microbiol. Suppl.* **27**, 117–127 (1987).
3. Fathepure, B. Z., Tiedje, J. M. & Boyd, S. A. Reductive dechlorination of hexachlorobenzene to tri- and dichlorobenzenes in anaerobic sewage sludge. *Appl. Environ. Microbiol.* **54**, 327–330 (1988).
4. Beurskens, J. E. M. et al. Dechlorination of chlorinated benzenes by an anaerobic microbial consortium that selectively mediates the thermodynamic most favorable reactions. *Environ. Sci. Technol.* **28**, 701–706 (1994).
5. Maymó-Gatell, X., Chien, Y. T., Gossett, J. M. & Zinder, S. H. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**, 1568–1571 (1997).
6. Holliger, C., Wohlfarth, G. & Diekert, G. Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiol. Rev.* **22**, 383–398 (1999).
7. Hugenholz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180**, 366–376 (1998).
8. von Wintzingerode, F., Selent, B., Hegemann, W. & Göbel, U. B. Phylogenetic analysis of an anaerobic, trichlorobenzene-transforming microbial consortium. *Appl. Environ. Microbiol.* **65**, 283–286 (1999).
9. Bosma, T. N. P., van der Meer, J. R., Schraa, G., Tros, M. E. & Zehnder, A. J. B. Reductive dechlorination of all trichloro- and dichlorobenzene isomers. *FEMS Microbiol. Ecol.* **53**, 223–229 (1988).
10. Holliger, C., Schraa, G., Stams, A. J. M. & Zehnder, A. J. B. Enrichment and properties of an anaerobic mixed culture reductively dechlorinating 1,2,3-trichlorobenzene to 1,3-dichlorobenzene. *Appl. Environ. Microbiol.* **58**, 1636–1644 (1992).
11. Middeldorp, P. J. M., de Wolf, J., Zehnder, A. J. B. & Schraa, G. Enrichment and properties of a 1,2,4-trichlorobenzene-dechlorinating methanogenic microbial consortium. *Appl. Environ. Microbiol.* **63**, 1225–1229 (1997).
12. Adrian, L., Manz, W., Szewzyk, U. & Görsch, H. Physiological characterization of a bacterial consortium reductively dechlorinating 1,2,3- and 1,2,4-trichlorobenzene. *Appl. Environ. Microbiol.* **64**, 496–503 (1998).
13. Adrian, L., Szewzyk, U. & Görsch, H. Bacterial growth linked to reductive dechlorination of trichlorobenzenes. *Biodegradation* (in press).
14. Giesbrecht, P., Kersten, T., Maidhof, H. & Wecke, J. Staphylococcal cell wall: morphogenesis and fatal variations in the presence of penicillin. *Microbiol. Mol. Biol. Rev.* **62**, 1371–1414 (1998).
15. Ludwig, W. et al. Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**, 554–568 (1998).
16. Pfennig, N. *Rhodocyclus purpureus* gen. nov. and sp. nov., a ring-shaped, vitamin B₁₂-requiring member of the family Rhodospirillaceae. *Int. J. Syst. Bacteriol.* **28**, 283–288 (1978).
17. Zehnder, A. J. B. & Wuhrmann, K. Titanium(III) citrate as a nontoxic oxidation-reduction buffering system for the culture of obligate anaerobes. *Science* **194**, 1165–1166 (1976).
18. Kalmbach, S., Manz, W. & Szewzyk, U. Isolation of new bacterial species from drinking water biofilms and proof of their *in situ* dominance with highly specific 16S rRNA probes. *Appl. Environ. Microbiol.* **63**, 4164–4170 (1997).
19. Bade, K., Manz, W. & Szewzyk, U. Behavior of sulfate reducing bacteria under oligotrophic conditions and oxygen stress in particle-free systems related to drinking water. *FEMS Microbiol. Ecol.* **32**, 215–233 (2000).

Acknowledgements

We thank T. Tran and P. Wendler for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to L.A. (e-mail: lorenz.adrian@tu-berlin.de). The 16S rRNA gene sequence of strain CBDB1 is deposited in GenBank under accession number AF230641.

Calcium stores regulate the polarity and input specificity of synaptic modification

Makoto Nishiyama^{*†‡}, Kyonsoo Hong^{*‡}, Katsuhiko Mikoshiba[§], Mu-ming Poo^{*‡} & Kunio Kato[§]

^{*} Department of Biology, University of California at San Diego, La Jolla, California 92093-0357, USA

[§] Mikoshiba Calciosignal Net Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Corporation (JST), Tokyo 113-0021, Japan

[†] These authors contributed equally to this work.

[‡] Present address: Department of Biochemistry, New York University School of Medicine, 550 First Avenue, New York, New York 10016, USA (M.N. and K.H.); Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA (M.-m.P.)

Activity-induced synaptic modification is essential for the development and plasticity of the nervous system^{1–3}. Repetitive correlated activation of pre- and postsynaptic neurons can induce persistent enhancement or decrement of synaptic efficacy, commonly referred to as long-term potentiation or depression^{2,3} (LTP or LTD). An important unresolved issue is whether and to what extent LTP and LTD are restricted to the activated synapses^{4–8}. Here we show that, in the CA1 region of the hippocampus, reduction of postsynaptic calcium influx by partial blockade of NMDA (*N*-methyl-D-aspartate) receptors results in a conversion of LTP to LTD and a loss of input specificity normally associated with LTP, with LTD appearing at heterosynaptic inputs. The induction of LTD at homo- and heterosynaptic sites requires functional ryanodine receptors and inositol triphosphate (InsP₃) receptors, respectively. Functional blockade or genetic

deletion of type 1 InsP₃ receptors led to a conversion of LTD to LTP and elimination of heterosynaptic LTD, whereas blocking ryanodine receptors eliminated only homosynaptic LTD. Thus, postsynaptic Ca²⁺, deriving from Ca²⁺ influx and differential release of Ca²⁺ from internal stores through ryanodine and InsP₃ receptors, regulates both the polarity and input specificity of activity-induced synaptic modification.

Spontaneous and induced theta and gamma oscillations have been implicated in regulating spatial memory in the hippocampus⁹. We first determined that correlated pre- and postsynaptic activation at theta frequency (5 Hz) can induce synaptic modifications in the CA1 region of rat hippocampal slices. Correlated activation comprises a train of stimuli at 5 Hz (for 16 s) delivered to one Schaffer collateral/commissural input, with each stimulus paired with postsynaptic injection of a spike-inducing depolarizing current (2 nA, 2 ms). When the onset of excitatory postsynaptic potentials (EPSPs) preceded the peak of postsynaptic action potentials by 5 ms ('positive' interval), the EPSCs of the stimulated ('homosynaptic') input showed a persistent increase in amplitude after the correlated activation, whereas that of unstimulated ('heterosynaptic' or control) input was not affected (Fig. 1a). When the onset of EPSPs was about 20 ms after ('negative' interval) postsynaptic spiking, the EPSC amplitude of both homo- and heterosynaptic inputs showed a persistent reduction following the correlated activation (Fig. 1b), suggesting induction of LTD in the stimulated pathway and the spread of LTD to the unstimulated pathway. We observed no synaptic modification when the interval between the pre- and postsynaptic activation was larger than 30 ms (Fig. 1c).

We further examined the dependence of synaptic modifications on the relative timing of pre- and postsynaptic activation. At the homosynaptic input, there is a 15-ms window for the induction of LTP and two distinct windows for the induction of LTD at –28 to –16 ms and at +15 to +20 ms (Fig. 2a). At heterosynaptic unstimulated inputs, we found no significant LTP, but two windows for LTD similar to that of the homosynaptic input (Fig. 2b, c). The potentiation window and the depression window at –28 to –16 ms are

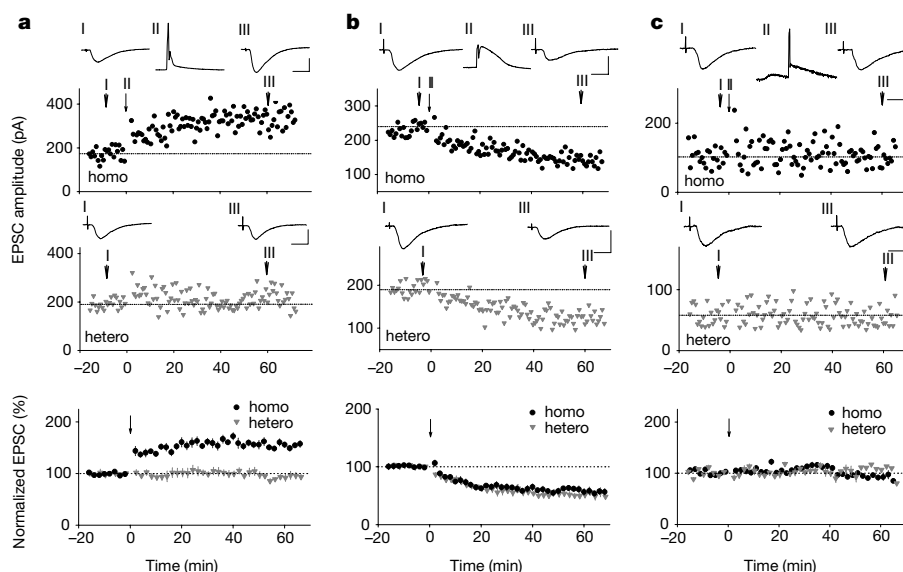


Figure 1 LTP/LTD induced by correlated pre- and postsynaptic activity in hippocampal CA1 pyramidal neurons. **a**, Top and middle, example of induction of LTP in the stimulated (homosynaptic) pathway and absence of LTP in the unstimulated (heterosynaptic) pathway. Data points represent the EPSC amplitude. Sample traces above: EPSCs (I, III) or membrane potential (II) at marked times. Postsynaptic spiking (5 Hz, 16 s) was elicited by depolarizing current pulses (2 nA, 2 ms) in current clamp ~5 ms after presynaptic field activation (5 Hz, 16 s). Scales: 200 pA (or 50 mV) and 20 ms. Bottom, summary of all results using correlated activation with positive intervals (+4 to +6 ms). Data represent mean EPSC amplitude ($\pm$ s.e.m., $n = 9$). The mean EPSC was significantly elevated over

the control ($t < 0$ min) at the homosynaptic ($t = 30$ –40 min, $P < 0.001$, Mann–Whitney *U*-test), but not the heterosynaptic pathway. **b**, As in **a**, except that the interval for pre- and postsynaptic activation was –28 to –16 ms (–22 ms in **b** top and middle). Significant reduction of EPSC amplitude was observed in both stimulated and unstimulated pathways ($n = 10$, $P < 0.001$, Mann–Whitney *U*-test) at $t = 30$ –40 min. Scales as in **a**. **c**, As in **a**, except that the interval for pre- and postsynaptic activation was < -30 ms or $> +30$ ms (44 ms in **c** top and middle). No significant change in EPSC amplitude was observed ($n = 6$). Scales: EPSCs, 200 (top) and 100 (middle) pA, 20 ms; potentials, 50 mV, 40 ms.

similar to those reported previously^{10,11}. The appearance of an additional depression window may result from a different spatio-temporal pattern of Ca^{2+} elevation or the presence of inhibitory inputs in the slice preparation. The lateral spread of depression at flanking windows may help to reduce the background noise and sharpen the effect of potentiation at selected pathways that are activated within the 15-ms 'potentiation window'. The time interval between the potentiation and depression is close to that of a gamma cycle, a feature that may be related to the idea that sequence recall in a place cell is linked to gamma cycles—on a theta cycle⁹.

To further explore the role of Ca^{2+} influx, we examined the effect of a partial blockade of NMDA subtype of glutamate receptors

(NMDARs) on synaptic changes induced by the correlated activity (at 5 Hz for 16 s) of positive interval of +4 to +6 ms ('LTP protocol'). Complete block of LTP/LTD was found at 5 and 50 μM AP-5 (D(-)-amino-5-phosphonovaleric acid) (Fig. 3a, b). At 1–3 μM AP5, LTP was reduced or even converted to LTD at the homosynaptic input, consistent with a previous report¹². Unexpectedly, we consistently observed LTD at the heterosynaptic input at these low AP5 concentrations (Fig. 3a, b). The extent of reduction in postsynaptic NMDAR activity was assayed by measuring NMDA components of EPSCs (Fig. 3c). Input specificity of LTP was lost when NMDAR activity was reduced by about 40% (at 1 μM AP5). These findings are consistent with the Bienenstock–Cooper–Munro (BCM) model

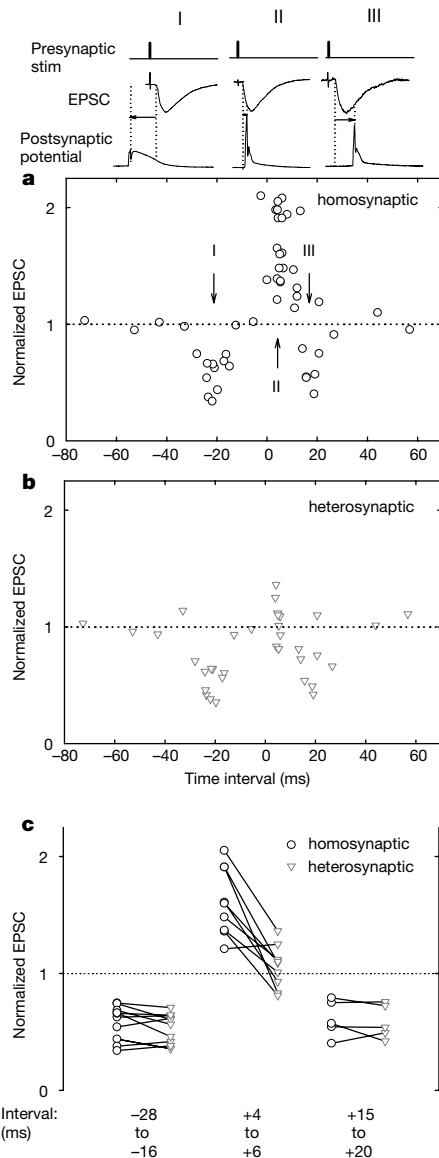


Figure 2 Critical time windows for the induction of LTP/LTD by correlated pre- and postsynaptic activation. Normalized mean EPSC amplitudes at 30–40 min after correlated activation are plotted against the interval between pre- and postsynaptic activation. The time interval refers to the time between the onset of the EPSC (top) and the peak of the postsynaptic action potential during each correlated activation. **a**, Summary of changes in the homosynaptic pathway ($n = 50$). **b**, Summary of changes in the heterosynaptic (unstimulated) pathway ($n = 34$). Data are from a subset of experiments such as those shown in **a**, in which the control pathway was monitored throughout the experiment. **c**, Changes of synaptic strength at the homo- and heterosynaptic pathways for experiments using time intervals of -28 to -16, +4 to +6 and +15 to +20 ms. Lines connect data points from the same experiment.

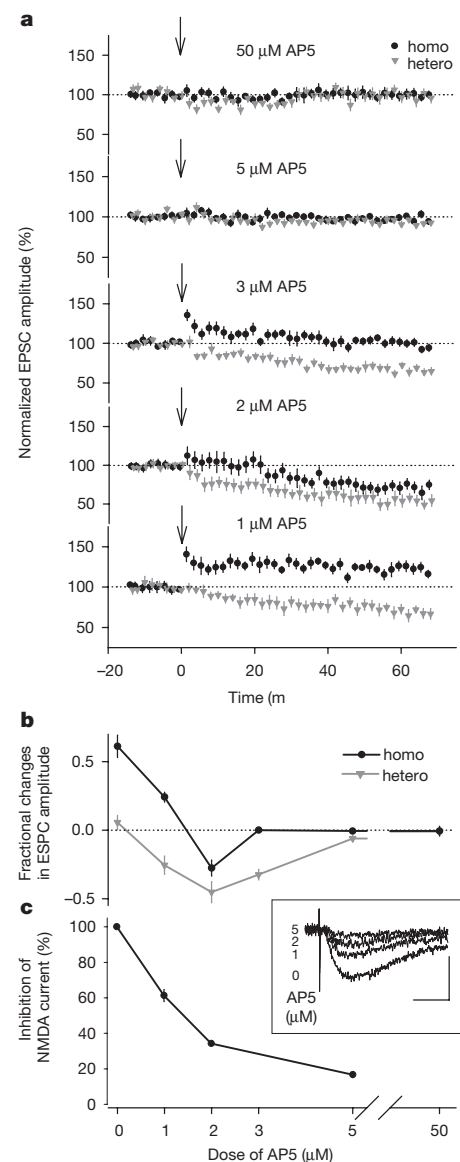


Figure 3 Effects of partial blockade of NMDA receptors. **a**, Experiments similar to that in Fig. 1a, except for the presence of AP5 (1–50 μM , $n = 5$ each). Correlated activity of positive intervals (+4 to +6 ms; 'LTP protocol') was applied at the time marked by the arrow. **b**, Dependence of synaptic modifications on the dose of AP5. The percentage change in the mean EPSC amplitude (at $t = 40$ –50 min) was plotted for homo- and heterosynaptic pathways (same data as in **a**). **c**, Percentage reduction of the NMDA component of EPSCs in CA1 pyramidal neurons after sequential application of AP5 from 0–5 μM (in CNQX and bicuculline, 20 μM each, with 1 mM QX-314 in patch pipettes; (s.e.m., $n = 5$). Data are normalized to that observed at 0 μM AP5. Stimulus strength is the same as in LTP experiments (100–200 pA in the absence of CNQX and bicuculline). Inset, sample traces of NMDA currents. Scales: 20 pA, 50 ms.

of synaptic modification^{13–15}, assuming that heterosynaptic spread of LTD is due to the spread of postsynaptic Ca^{2+} elevation from homo- to heterosynaptic sites.

Release of Ca^{2+} from internal stores, triggered either by postsynaptic Ca^{2+} influx¹⁶ or by activation of metabotropic glutamate receptors (mGluRs)¹⁷ is involved in activity-induced synaptic modification. We thus examined the effect of a monoclonal antibody (18A10) against the type 1 InsP_3 receptor (InsP_3R), a predominant subtype of InsP_3Rs found in CA1 pyramidal neurons and localized primarily in the dendritic shaft and soma¹⁸. The function-blocking activity of 18A10 was shown by its blocking effect on InsP_3 -induced Ca^{2+} release from hippocampal microsomal fractions (see Supplementary Information). After postsynaptic loading of 18A10 ($160\text{ }\mu\text{g ml}^{-1}$), repetitive correlated activation with an interval of -24 to -20 ms ('LTD protocol') resulted in homosynaptic LTP, without significant heterosynaptic effect (Fig. 4a). Experiments in which control immunoglobulin- γ (IgG; $160\text{ }\mu\text{g ml}^{-1}$) was used instead of 18A10 showed normal homo- and heterosynaptic LTD. Loading of 18A10 did not block the induction of LTP by the LTP protocol, but increased the extent of LTP (Fig. 5a). Unexpectedly, in 3 out of 6 cases, significant LTP was found also in the heterosynaptic pathway (Fig. 5a), indicating that there may be a loss of input specificity. Furthermore, using slices obtained from InsP_3R -deficient mice¹⁹, we found that LTD protocol resulted in a slight LTP at the homosynaptic input and no synaptic change at the heterosynaptic site (Fig. 4b), whereas LTP protocol resulted in an elevated LTP at the homosynaptic input and a slight potentiation at the heterosynaptic input (Fig. 5b). Parallel recordings using slices from wild-type mice (Figs 4b, 5b) yielded results identical to those obtained from rat slices (Fig. 1a, b). Together, these results indicate that InsP_3R activity determines the polarity and the extent of synaptic changes and is responsible for the heterosynaptic spread of LTD. The finding that heterosynaptic LTD normally triggered by tetanic stimulation in the CA1 area of hippocampus of the wild-type mice was completely absent in the type-1 InsP_3R deficient mice (K.K., unpublished data) is also consistent with this idea.

Metabotropic GluR-dependent LTD may be induced at Schaffer collateral-CA1 synapses²⁰. One of the downstream pathways of mGluR activation is the production of InsP_3 and the resultant

Ca^{2+} release through InsP_3Rs ²¹. After bath-application of an antagonist of mGluRs, (*S*)- α -methyl-4-carboxyphenylglycine (MCPG, 1 mM), stimulation of Schaffer collateral/commissural-CA1 synapses with the LTD protocol induced significant LTP in both homo- and heterosynaptic pathways, with the mean EPSC amplitude elevated to 142 ± 10 and $133 \pm 13\%$ (s.e.m., $n = 7$) of the control level, respectively, similar to that induced by antibody blockade or genetic deletion of InsP_3Rs . This is consistent with a MCPG-induced reduction of InsP_3R activity, although the effect of MCPG on presynaptic mGluRs and other downstream effectors²⁰ cannot be excluded.

The above studies showed that reduction of functional InsP_3Rs resulted in marked changes in the polarity and extent of synaptic modification: correlated activation of negative intervals led to LTP instead of LTD, whereas that of positive intervals led to a larger LTP. What is the cellular mechanisms underlying these changes? We found that the resting potential, membrane resistance, time constant, firing threshold, and the height and half-width of action potentials were not significantly different between the type 1 InsP_3R -deficient and wild-type mice (data not shown). However, we observed a marked reduction of spike frequency adaptation²² in response to a long depolarizing step in mutant slices (see Supplementary Information), suggesting reduced K^+ channel activity in mutant mice. Ca^{2+} -activated K^+ channels can be activated through Ca^{2+} released from internal stores²³. Dysfunction of InsP_3Rs , therefore, may cause the inactivation of Ca^{2+} -activated K^+ channels, resulting in an increased postsynaptic excitability. We also found that synaptic NMDA currents near the resting membrane potential, as recorded in the presence of 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) and bicuculline, were significantly higher in mutant slices or in wild-type slices after loading of the specific type-1 InsP_3R antibody 18A10 (see Supplementary Information), suggesting a reduced blockade of NMDARs by Mg^{2+} . Thus, a higher level of local Ca^{2+} influx during correlated activation may be responsible for the observed changes in synaptic modification.

In addition to Ca^{2+} release mediated by InsP_3Rs , internal Ca^{2+} release through RyRs is also involved in the induction of LTD by a low frequency stimulation²⁴. Here we found postsynaptic loading with ryanodine ($100\text{ }\mu\text{M}$), which blocks ryanodine receptors (RyRs) in the internal stores, resulted in a failure of standard LTD

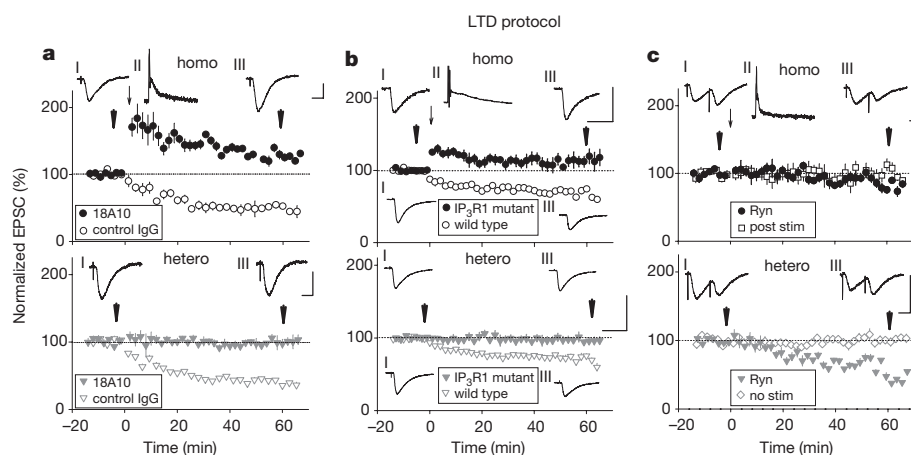


Figure 4 Effects of blocking Ca^{2+} release from internal stores on the induction of LTD. **a**, Postsynaptic neurons were loaded with a function-blocking antibody (18A10, $160\text{ }\mu\text{g ml}^{-1}$; $n = 6$) against the type 1 InsP_3R or control IgG ($160\text{ }\mu\text{g ml}^{-1}$; $n = 3$). For 18A10 loading, homosynaptic LTP instead of LTD (upper graph) and no heterosynaptic change (lower graph) were observed. For control IgG loading, significant LTD was observed in both pathways. Scales: 100 pA (30 mV), 20 ms . **b**, LTD protocol applied to slices obtained from InsP_3R -deficient mice ($\text{IP}_3\text{R1}$ mutant; $n = 6$) resulted in a slight LTP at the homosynaptic input (upper graph) and no heterosynaptic change (lower graph). The induction protocol was the same as that used for rat slices (Fig. 1b), except that the initial EPSC amplitude was in the range of 50 – 70 pA . Control experiments using slices from

wild-type mice ($n = 6$) showed significant LTD in both homo- and heterosynaptic pathways. Scales: 100 pA (or 200 mV), 50 ms . **c**, LTD protocol applied to slices with postsynaptic loading of ryanodine (Ryn; $100\text{ }\mu\text{M}$) resulted in no synaptic change at the homosynaptic pathway (upper graph) but LTD at heterosynaptic pathway (lower graph). Samples of paired EPSCs above show responses induced by sequential stimulation (interval 50 ms) of the homo- then heterosynaptic pathways (upper graph), and the hetero- then homosynaptic pathways (lower graph), respectively. Scales: 200 pA (or 100 mV), 20 ms . Control experiments: post stim, postsynaptic stimulation alone ($n = 6$); no stim, no pre- or postsynaptic activation ($n = 4$).

protocol in inducing homosynaptic LTD (Fig. 4c). Surprisingly, LTD was observed at the heterosynaptic input. These results were not due to the loading of ryanodine, as, for neurons loaded with ryanodine, repetitive postsynaptic activation alone ('post stim') or in the absence of any stimulation ('no stim') resulted in no change in synaptic efficacy (Fig. 4c). Thus, Ca^{2+} release from RyRs is essential for the induction of homosynaptic LTD by the correlated activity. Furthermore, after postsynaptic loading of ryanodine, standard LTP protocol induced an elevated level of homosynaptic LTP and a significant heterosynaptic LTD (Fig. 5c).

From our results, we propose the following model for the interplay of localized Ca^{2+} influx and Ca^{2+} release from internal stores (Fig. 6a). Ca^{2+} influx through NMDARs and voltage-dependent Ca^{2+} channels, together with mGluR activation, triggers InsP_3 -dependent Ca^{2+} release from internal stores, which in turn causes further Ca^{2+} release through RyRs^{17,25}. When a high-level of Ca^{2+}

transient is created, for example, under conditions that induce LTP, InsP_3 Rs become desensitized²⁶. Furthermore, Ca^{2+} release from internal stores may reduce postsynaptic neuronal excitability through Ca^{2+} -activated K^+ channels²³ and activated InsP_3 Rs may also reduce NMDAR activity through an unknown mechanism. Consistent with the idea that the polarity of homosynaptic modification depends on the level of postsynaptic Ca^{2+} —a high-level Ca^{2+} triggers LTP, whereas a modest-level Ca^{2+} triggers LTD (see refs 27, 28)—reduction of NMDAR activity (which reduces Ca^{2+} influx) results in a reduced LTP or even a conversion to LTD (Fig. 3). Inhibition or deletion of InsP_3 Rs elevates neuronal excitability and NMDAR activity, resulting in a higher Ca^{2+} level and thus elevated LTP (Figs 4b, 5b). Inhibition of RyRs may also result in a higher Ca^{2+} , through reduced activation of K^+ channels²⁵, leading to the absence of homosynaptic LTD following LTD protocol, but an increased homosynaptic LTP following LTP protocol.

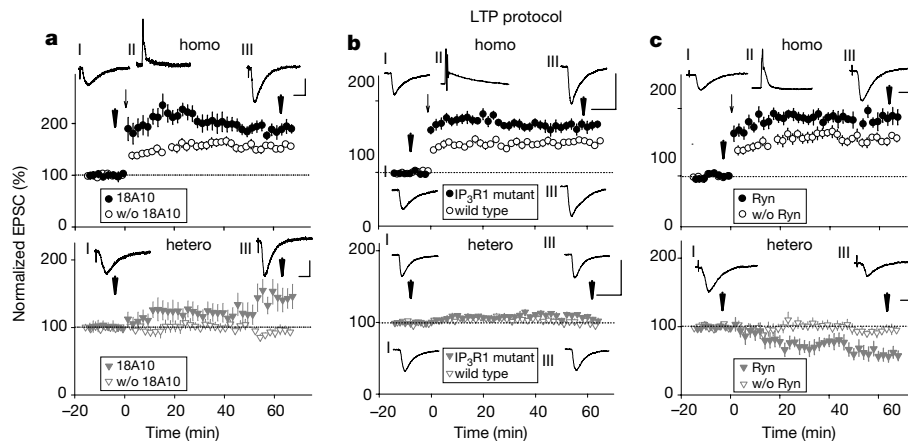


Figure 5 Effects of blocking internal Ca^{2+} release on the induction of LTP. **a**, As in Fig. 4a, except that LTP protocol was used (with 18A10 loading, $n = 6$). The extent of potentiation at $t = 30\text{--}40$ min was significantly greater ($P < 0.03$, Mann–Whitney U -test) than that found in the absence of antibody loading (w/o 18A10, same data as in Fig. 1a, bottom). Slight potentiation was also found in the heterosynaptic pathway. Scales: 100 pA (or 30 mV), 20 ms. **b**, As in Fig. 4b, except that LTP protocol was used ($n = 6$). Significant enhancement of homosynaptic LTP was observed ($164 \pm 8\%$ and $142 \pm 2\%$ of the

control for mutant and wild-type mice at $t = 30\text{--}40$ min, respectively, $P < 0.05$, t -test). No significant synaptic change was observed for the heterosynaptic input ($n = 6$, $P = 0.077$, t -test). **c**, Postsynaptic loading of ryanodine resulted in a greater extent of homosynaptic LTP after the standard LTP induction protocol ($n = 5$, $P < 0.03$, Mann–Whitney U -test) than that found in the absence of ryanodine (same data set as in Fig. 1, top). Marked depression was found in the heterosynaptic pathway. Scales as in **a**.

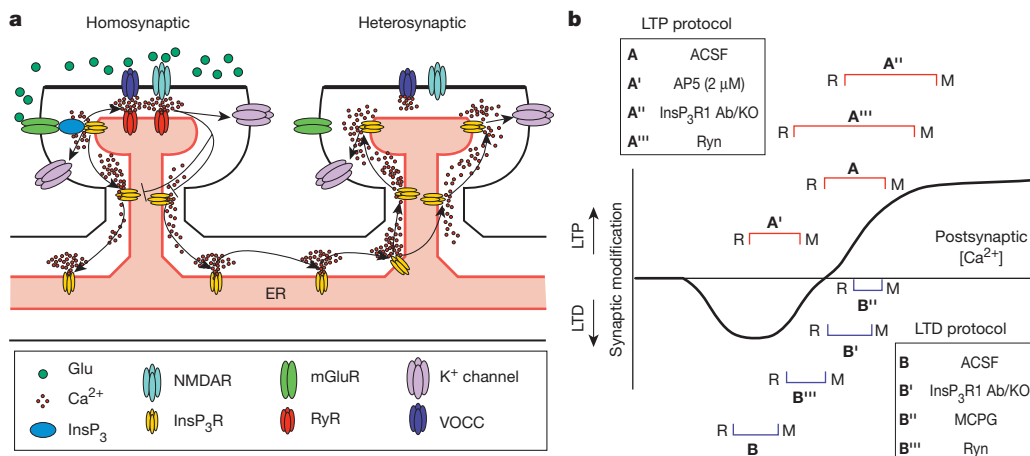


Figure 6 A model for the roles of postsynaptic Ca^{2+} signalling in synaptic modification. **a**, Postsynaptic Ca^{2+} regulation. Influx of Ca^{2+} through NMDARs and voltage-operated Ca^{2+} channels (VOCCs), together with mGluR-dependent Ca^{2+} release from InsP_3 Rs, triggers further Ca^{2+} release from RyRs and InsP_3 R-dependent propagation of Ca^{2+} waves in the dendrite. Localized high-level Ca^{2+} transients in the spine may desensitize InsP_3 Rs, preventing propagation of Ca^{2+} waves. Release of Ca^{2+} from internal stores also activates plasma membrane K^+ channels. **b**, Dependence of synaptic modification on postsynaptic

Ca^{2+} . The LTP and LTD induced by correlated activity of positive and negative intervals, respectively, at homosynaptic (M) and heterosynaptic (R) pathways are indicated by the bracket. For different experimental conditions, the polarity, degree, and input-specificity of synaptic modification are shifted in accord with the prediction of the BCM model (solid curve), assuming postsynaptic Ca^{2+} levels to be the determining variable. A–A'' and B–B'' (boxed) denote different experimental conditions. ACSF, artificial cerebrospinal fluid; Ab, antibody; KO, knock-out.

To account for the lack of input specificity for LTD and the dependence of input specificity on postsynaptic Ca^{2+} , we propose that synaptic activation by correlated activity results in Ca^{2+} elevation that spreads from the active synapses to distant unstimulated heterosynaptic sites by propagating Ca^{2+} waves²⁹ mediated by InsP_3 Rs, which are localized primarily to the soma and dendritic shafts¹⁸. As illustrated by the BCM model for the induction of LTP/LTD^{13–15} (Fig. 6b), standard LTP protocol (A) results in a high level of Ca^{2+} and LTP only at the homosynaptic input. No change is induced at the heterosynaptic site, because high-level Ca^{2+} elevation at the homosynaptic site had caused RyR-dependent desensitization of InsP_3 Rs. Standard LTD protocol (B) results in a modest elevation of Ca^{2+} that leads to homo- and heterosynaptic LTD, the latter owing to InsP_3 R-mediated long-range spread of Ca^{2+} . For modification induced by the LTP protocol, partial reduction of NMDAR-mediated Ca^{2+} influx results in a downward slide and LTD at both inputs (A'). Reduction of functional InsP_3 Rs led to an upward slide, with larger homosynaptic LTP and a loss of input specificity (A''). For the LTD protocol, an upward slide occurs after reduction of functional InsP_3 Rs by antibody blockade, genetic deletion (B'), or by MCPG (B''), leading to a conversion of LTD to LTP at the homosynaptic input and corresponding changes at the heterosynaptic site. Although blocking RyRs by ryanodine resulted in a higher level of local Ca^{2+} , and so upward shift of homosynaptic modification (A''' and B'''), the lack of RyR-dependent desensitization of InsP_3 Rs allowed InsP_3 R-dependent spread of LTD to heterosynaptic sites. By this account, input specificity in synaptic modification is not an intrinsic property associated with LTP or LTD, but a dynamic variable linked to the pattern of long-range Ca^{2+} signalling in the postsynaptic dendrite.

We have shown three important features of bi-directional synaptic modification induced by correlated activity in the CA1 region of the hippocampus. First, correlated activity of the same low frequency of 5 Hz can induce both LTP and LTD, depending on the precise timing of pre- and postsynaptic activation, suggesting a synaptic mechanism for processing and storage of information with a time resolution in the order of 10 ms. Second, the polarity and magnitude of synaptic changes are readily modified by manipulations of Ca^{2+} influx or Ca^{2+} release from internal stores in the postsynaptic neuron. Third, the input specificity depends on the pattern of postsynaptic Ca^{2+} and InsP_3 R-mediated Ca^{2+} release is essential for the long-range propagation of LTD. Together, these results underscore the importance of spatiotemporal patterns of postsynaptic Ca^{2+} , for which the differential dendritic localization of InsP_3 - and ryanodine-sensitive stores is critical. □

Methods

Hippocampal slice preparation

We prepared hippocampal slices according to the standard procedure³⁰. Male albino rats (26–33 days), mice lacking type I InsP_3 Rs and wild-type littermates (20–23 days) were anaesthetized with halothane and decapitated using a guillotine. Mouse genotypes were identified using a described procedure¹⁹. Hippocampi were dissected rapidly and placed in gassed (95% O_2 / 5% CO_2) extracellular solution at 10 °C containing (in mM): 124 NaCl, 3 KCl, 2.6 CaCl_2 (2.0 for mice), 1.3 MgSO_4 (2.0 for mice), 1.25 NaH_2PO_4 , 22 NaHCO_3 and 10 glucose. Transverse slices (500- μm thick) were cut with a rotary tissue slicer (DTY7700) and maintained in an incubation chamber for at least 2 h at room temperature. For experiments, individual slices were transferred to a submersion recording chamber and perfused continuously with extracellular solution (~ 4.0 – 4.5 ml min^{-1}) at room temperature (23–26 °C).

Whole-cell recordings

Whole-cell recordings were made in the cell body layer of CA1 by the 'blind' patch-clamp method, using an Axopatch 200B amplifier (Axon Inst.). Stimuli were applied at 0.05 Hz and alternated between two non-overlapping Schaffer collateral/commissural inputs, using bipolar electrodes (MCE-100, RMI). The non-overlapping nature of the two inputs was confirmed by applying paired-pulse test stimuli (with 50-ms intervals) alternately to show the absence of cross facilitation between the two inputs (see traces in Fig. 4c). Constant current pulses (amplitude, 6–14 μA ; duration, 300 μs) normally evoked EPSCs with amplitudes in the range of 100–200 pA. We considered only data with initial EPSC amplitude > 70 pA, as synaptic modification induced by correlated activity in these slices

differs significantly between inputs with strength more than and less than 70 pA (M.N., et al., unpublished data). Data were filtered at 2 kHz and digitized at 10 kHz. Patch electrodes were pulled from borosilicate glass (1.2-mm optical density) and had a resistance of ~ 3.5 – $6 \text{ M}\Omega$. Pipettes were normally filled with a solution containing (in mM): 130 caesium methanesulphonate, 10 tetraethylammonium chloride, 5 NaCl, 0.25 1,2-bis(2-aminophenoxy)ethane-N,N',N'-tetraacetic acid (BAPTA), 10 HEPES, 4 Mg-ATP, with pH adjusted to 7.35 (using CsOH). The series resistance was typically 10–14 $\text{M}\Omega$ and was compensated 50–80% during the experiment. Drugs were purchased from RBI (AP5), Calbiochem (ryanodine) or Tocris Cookson (MCPG). For intracellular loading of ryanodine and anti- InsP_3 R antibody, correlated activation was applied 15–20 min after break in.

Received 1 June; accepted 7 September 2000.

- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Martin, S. J., Grimwood, P. D. & Morris, R. G. M. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu. Rev. Neurosci.* **23**, 649–711 (2000).
- Schuman, E. M. & Madison, D. V. Locally distributed synaptic potentiation in the hippocampus. *Science* **263**, 532–536 (1994).
- Staubli, U. V. & Ji, Z.-X. The induction of homo- vs. heterosynaptic LTD in area CA1 of hippocampal slices from adult rats. *Brain Res.* **714**, 169–176 (1996).
- Engert, F. & Bonhoeffer, T. Synapse specificity of long-term potentiation breaks down at short distances. *Nature* **388**, 279–284 (1997).
- Fitzsimonds, R. M., Song, H.-j. & Poo, M.-m. Propagation of activity-dependent synaptic depression in simple neural networks. *Nature* **388**, 439–448 (1997).
- Tao, H.-z., Zhang, L. I., Bi, G.-q. & Poo, M.-m. Selective presynaptic propagation of long-term potentiation in defined neural networks. *J. Neurosci.* **20**, 3233–3243 (2000).
- Lisman, J. E. Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate–CA3 interactions. *Neuron* **22**, 233–242 (1999).
- Zhang, L. I. et al. A critical window for cooperation and competition among developing retinotectal synapses. *Nature* **395**, 37–44 (1998).
- Bi, G.-q. & Poo, M.-m. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- Cummings, J. A., Mulkey, R. M., Nicoll, R. A. & Malenka, R. C. Ca^{2+} signaling requirements for long-term depression in the hippocampus. *Neuron* **16**, 825–833 (1996).
- Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
- Artola, A. & Singer, W. Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* **16**, 480–487 (1993).
- Bear, M. Mechanism for a sliding synaptic modification threshold. *Neuron* **15**, 1–4 (1995).
- Emptage, N., Bliss, T. V. P. & Fine, A. A single synaptic event evokes NMDA receptor-mediated release of calcium from internal stores in hippocampal dendritic spines. *Neuron* **22**, 115–124 (1999).
- Berridge, M. J. Neuronal calcium signaling. *Neuron* **21**, 13–26 (1998).
- Furuichi, T., Kohda, K., Miyawaki, A. & Mikoshiba, K. Intracellular channels. *Curr. Opin. Neurobiol.* **4**, 294–303 (1994).
- Matsumoto, M. et al. Ataxia and epileptic seizures in mice lacking type I inositol 1,4,5-trisphosphate receptor. *Nature* **379**, 168–171 (1996).
- Oliet, S. H. R., Malenka, R. C. & Nicoll, R. A. Two distinct forms of long-term depression coexist in CA1 hippocampal pyramidal cells. *Neuron* **18**, 969–982 (1997).
- Pin, J.-P. & Duvoisin, R. Neurotransmitter receptors. I. The metabotropic glutamate receptors: structure and functions. *Neuropharmacology* **34**, 1–26 (1995).
- Conner, J. A. & Stevens, C. F. Prediction of repetitive firing behaviour from voltage clamp data on an isolated neurone soma. *J. Physiol.* **213**, 31–53 (1971).
- Cordoba-Rodriguez, R., Moore, K. Y., Kao, J. P. Y. & Weinreich, D. Calcium regulation of a slow post-spike hyperpolarization in vagal afferent neurons. *Proc. Natl Acad. Sci. USA* **96**, 7650–7657 (1999).
- Futatsugi, A. et al. Facilitation of NMDA-independent LTP and spatial learning in mutant mice lacking ryanodine receptor type 3. *Neuron* **24**, 701–713 (1999).
- Fagni, L., Chavis, P., Ango, F. & Bockaert, J. Complex interactions between mGluRs, intracellular Ca^{2+} stores and ion channels in neurons. *Trends Neurosci.* **23**, 80–88 (2000).
- Bezprozvanny, I., Watras, J. & Ehrlich, B. E. Bell-shaped calcium-response curves of $\text{Ins}(1,4,5)\text{P}_3$ - and calcium-gated channels from endoplasmic reticulum of cerebellum. *Nature* **351**, 751–754 (1991).
- Lisman, J. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proc. Natl Acad. Sci. USA* **86**, 9574–9578 (1989).
- Yang, S.-N., Tang, Y.-G. & Zucker, R. S. Selective induction of LTP and LTD by postsynaptic $[\text{Ca}^{2+}]_i$ elevation. *J. Neurophysiol.* **81**, 781–787 (1999).
- Nakamura, T., Barbara, J.-G., Nakamura, K. & Ross, W. N. Synergistic release of Ca^{2+} from InsP_3 -sensitive stores evoked by synaptic activation of mGluRs paired with backpropagating action potentials. *Neuron* **24**, 727–737 (1999).
- Kato, K., Clifford, D. B. & Zorumski, C. F. Long-term potentiation during whole-cell recording in rat hippocampal slices. *Neuroscience* **53**, 39–47 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Michikawa for providing an antibody against InsP_3R_1 ; T. V. P. Bliss and R. C. Malenka for helpful discussions and suggestions; and N. Spitzer, J. R. Henley, D. Zacharias, A. F. Schinder, S. Andersen and F. Engert for critical comments on the manuscript. This work was supported in part by a grant from USNIH.

Correspondence and requests for materials should be addressed to M.-m.P. (e-mail: mpoo@uclink.berkeley.edu).

Analysis of calcium channels in single spines using optical fluctuation analysis

Bernardo L. Sabatini & Karel Svoboda

Howard Hughes Medical Institute, Cold Spring Harbor Laboratory,
1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

Most synapses form on small, specialized postsynaptic structures known as dendritic spines¹. The influx of Ca^{2+} ions into such spines—through synaptic receptors and voltage-sensitive Ca^{2+} channels (VSCCs)—triggers diverse processes that underlie synaptic plasticity². Using two-photon laser scanning microscopy³, we imaged action-potential-induced transient changes in Ca^{2+} concentration in spines and dendrites of CA1 pyramidal neurons in rat hippocampal slices⁴. Through analysis of the large trial-to-trial fluctuations in these transients, we have determined the number and properties of VSCCs in single spines. Here we report that each spine contains 1–20 VSCCs, and that this number increases with spine volume. We are able to detect the opening of a single VSCC on a spine. In spines located on the proximal dendritic tree, VSCCs normally open with high probability (~ 0.5) following dendritic action potentials. Activation of GABA_B receptors reduced this probability in apical spines to ~ 0.3 but had no effect on VSCCs in dendrites or basal spines. Our studies show that the spatial distribution of VSCC subtypes and their modulatory potential is regulated with submicrometre precision.

We imaged spines and dendrites of CA1 pyramidal neurons filled with a Ca^{2+} -sensitive fluorophore (Fig. 1). Action potentials propagate into these dendrites and trigger Ca^{2+} influx ($\Delta[\text{Ca}]_{\text{MP}}$) by opening VSCCs⁵. We measured action-potential-induced fluorescence transients simultaneously in spine heads and in small dendrites, under conditions where changes in fluorescence are proportional to changes in the concentration of intracellular free Ca^{2+} , $\Delta[\text{Ca}^{2+}]$. The tortuous spine neck serves to diffusionally isolate spine heads from their parent dendrites over timescales of 20–100 ms (ref. 6). Therefore, immediately after an action potential, $\Delta[\text{Ca}^{2+}]$ in spines will be due exclusively to opening of VSCCs on the spine head, with no contribution from dendritic VSCCs.

Following an action potential, $\Delta[\text{Ca}^{2+}]$ rose quickly (< 2 ms), showing that VSCCs exist in spine heads and dendrites⁴ (Fig. 1b). We determined the types of VSCCs present by the application of pharmacological blockers^{7,8} (Fig. 1c). ω -Conotoxin-MVIIIC (CTX), a blocker of N/P/Q-type VSCCs, had no effect on $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ (dendrites: $95 \pm 5\%$ of control, spines: $96 \pm 8\%$, $n = 4$). CTX diffused into the slice and blocked presynaptic Ca^{2+} influx, as it reduced synaptic potentials by more than 90% in these same cells (EPSP, Fig. 1c). Blockade of L-type channels by nimodipine reduced $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ in dendrites but not spines (dendrites: $86 \pm 8\%$ of control, spines: $99 \pm 4\%$, $n = 5$). Blockade of NMDA-Rs had no effect on $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ (dendrites: $95 \pm 3\%$ of control, spines: $97 \pm 3\%$ in spines, $n = 4$). Because blocking L-, N-, P- and Q-type VSCCs did not affect $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ in the spine, and low-voltage-activated channels are inactivated at resting membrane potentials of our experiments (-60 to -68 mV)⁹, action-potential-evoked influx of Ca^{2+} in spines is probably primarily through R-type channels.

Because the areas of spine membranes are small ($< 0.5 \mu\text{m}^2$)¹, spines are expected to contain fewer VSCCs and to show larger fluctuations in $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ than dendrites. Larger trial-to-trial fluctuations in fluorescence were observed in spines than on their parent dendrites (Fig. 1d). Fluctuations were stationary and subsequent trials were uncorrelated (Fig. 2a), indicating that the number of VSCCs does not vary during the experiment. As the

low mobility of membrane proteins¹⁰ prevents diffusion of VSCCs into and out of spines during an interstimulus interval (5 s), our measurements have sufficient time resolution to detect changes in the number of channels if they occur. We analysed trial-to-trial fluctuations further to extract the number and opening probability of VSCCs in individual spines. Dark noise (before shutter opening), baseline fluorescence (between shutter opening and action potential initiation), and response fluorescence (after action potential initiation) were measured during each trial (Fig. 2b, c). We selected for analysis pairs of spine heads and parent dendrites whose fluctuations in baseline fluorescence were no greater than expected from photon shot noise and dark noise. In these structures, signals were not contaminated by noise sources such as movement of the spine, fluorophore bleaching, photodamage, and run-down of calcium currents. Furthermore, fluctuations in resting $[\text{Ca}^{2+}]$ due to spontaneous VSCC openings were negligible. Following a back-propagating action potential, the variance of the fluorescence increased and was larger than expected from shot noise (Fig. 2b, c). The additional variance was caused by stochastic opening of VSCCs in the spine and dendrite.

In a spine containing N channels that open independently with probability p per action potential, the number of channels opened by an action potential is governed by the binomial distribution¹¹.

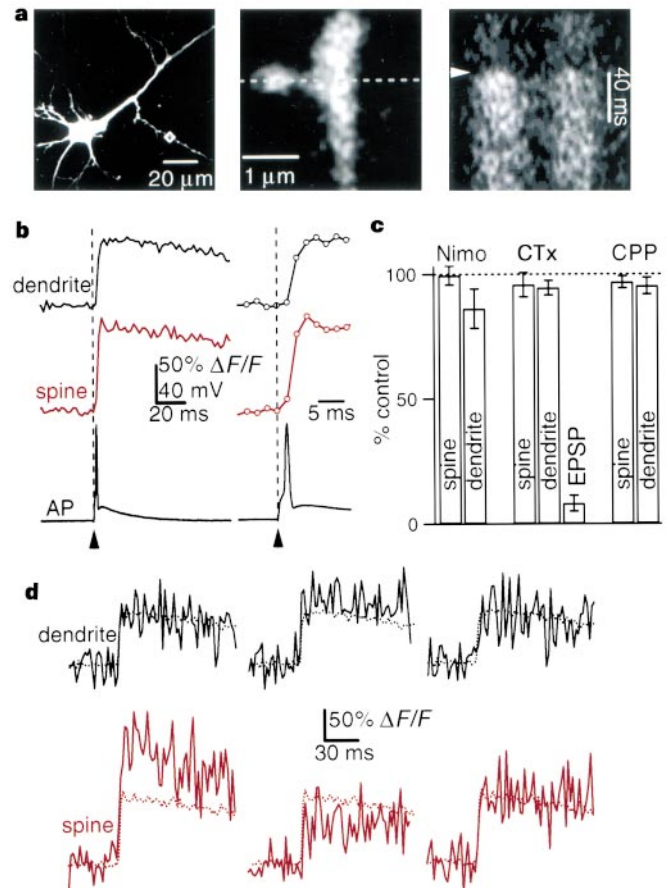


Figure 1 Action-potential-evoked Ca^{2+} influx in spines and dendrites. **a**, Neuron filled with Fluo-4 (left), and magnified view of the region outlined by the box showing an apical spine and its parent dendrite (middle). Right, line scan over the spine head and neighbouring dendrite. Arrowhead, time when action potential was stimulated. **b**, Average of 57 fluorescence transients in the spine and dendrite, aligned with the somatic action potential (AP). Dashed line, start of somatic current injection. Right, rising phase of the fluorescence transients indicating data points ($\Delta t = 2$ ms). **c**, Effects of blockers on action-potential-evoked fluorescence transients. The effect of CTX on synaptic potentials (EPSP) is also shown. **d**, Action-potential-evoked fluorescence transients (solid lines) superimposed on the average (dashed lines).

The coefficient of variation of the number of open channels, $CV_{\text{open}} = \sqrt{(1-p)/Np}$, is proportional to the measured coefficient of variation of $\Delta[\text{Ca}^{2+}]$ ($CV_{F\Delta[\text{Ca}]}$) such that $CV_{\text{open}}^2 = kCV_{F\Delta[\text{Ca}]}^2$. k quantifies the fraction of the variance in $\Delta[\text{Ca}^{2+}]$ due to fluctuations in Ca^{2+} influx through open channels, and can be estimated as $0.8 \leq k \leq 1$ (see Methods). In some spines action potentials occasionally failed to trigger a rapid rise in fluorescence (Fig. 3b–f). These failures were not caused by lack of action potential propagation because $\Delta F/F$ in the dendrite was independent of scoring the spine signal as failure or success (Fig. 3g). The probability of failure is $P(0) = (1-p)^N$; this second equation allowed us to calculate p and N (Fig. 3f, see Methods). In seven spines that showed failures in more than 5% of trials, we found $N = 5 \pm 0.8$ (range 3–7) and $p = 0.52 \pm 0.05$ (range 0.37–0.75). Spines that showed fewer failures were not used to calculate p as the scarcity of failures did not allow for reliable estimation of $P(0)$.

We investigated whether activation of G-protein-coupled receptors modulates the properties of VSCCs in dendritic spines. Bath application of the GABA_B agonist baclofen significantly reduced $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ in apical spines ($56 \pm 7\%$ of control, $n = 5$) but not in the neighbouring dendrite ($97 \pm 4\%$ of control, $n = 10$) (Fig. 4a) or basal spines ($99 \pm 3\%$ of control, $n = 5$). The effects of baclofen were reversed upon washout (2/2, not shown). Reduced $\Delta[\text{Ca}^{2+}]$ in apical spines was not due to alterations of the shape or reliability of propagation of the action potential, as baclofen had no effect on

$\Delta F/F$ in the neighbouring dendrite which is isopotential with the spine¹². The fraction of spines with failures was higher in baclofen than in control conditions (40% ($n = 10$) versus 12% ($n = 55$)), indicating that fewer channels opened in response to an action potential (Fig. 4b–d). In this fraction $p = 0.31 \pm 0.04$ (range 0.25–0.41, $n = 4$), which is significantly lower than the value found in control conditions (Fig. 4e). As expected from a reduction in p , baclofen application increased $CV_{F\Delta[\text{Ca}]}$ in apical spines ($160 \pm 8\%$ of control, $n = 5$, Fig. 4f–h). Lowering extracellular $[\text{Ca}^{2+}]$, which reduces Ca^{2+} influx per open channel and thus is equivalent to modulation of single channel conductance, reduced $\Delta[\text{Ca}^{2+}]$ in all spines ($\Delta F/F = 63 \pm 6\%$ of control, $n = 4$) but had no effect on $CV_{F\Delta[\text{Ca}]}$ ($101 \pm 4\%$ of control, $n = 4$) (Fig. 4h).

Using $p = 0.52$, obtained from the analysis of spines that showed failures in control conditions (Fig. 3), we calculated the number of VSCCs that open following an action potential in spines that did not show failures. We find $Np = 3.8 \pm 0.3$ (range 0.4–10.4, $n = 55$). Approximately half of this range can be explained by spine size: the number of open channels is proportional to spine volume with a slope of 120 channels μm^{-3} (Pearson's $R = 0.75$, $P < 0.05$) (Fig. 5).

In summary, we find that more than half of the VSCCs on each spine open following an action potential. Electrical recordings have previously described an enrichment of R-type channels in hippocampal dendrites; however, L-, N- and P/Q-type VSCCs were also found^{9,13,14}. We find that action-potential-evoked Ca^{2+} influx is

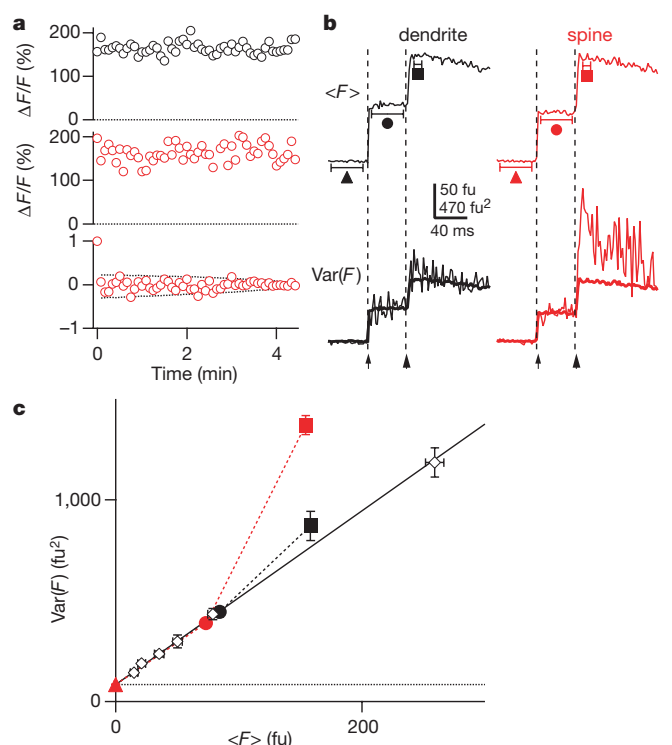


Figure 2 Variance analysis. **a**, Amplitudes of fluorescence transients in the dendrite (top) and spine (middle) shown in Fig. 1. Bottom, autocorrelation of response amplitudes in the spine (circles). Dashed lines, ± 2 s.d. of shuffled data sets. **b**, Mean fluorescence ($\langle F \rangle$) and variance of fluorescence ($\text{var}(F)$) in the dendrite (black) and spine (red). Top: horizontal bars show duration of measurements plotted in **c**. Bottom: thick trace, variance expected from photon shot noise and dark noise; vertical dashed lines, opening of the illumination shutter (arrow) and somatic current injection (arrowhead). **c**, Variance plotted against mean for fluorescence from the spine (red) and dendrite (black) with the shutter closed (triangles), before action potential stimulation (circles), and in the 10 ms following the peak of the fluorescence transient (squares) as indicated in panel **b**. Rhomboids, variance/mean relationship obtained by imaging the patch pipette while varying illumination intensity. Horizontal dotted line, variance due to dark noise; solid line, variance expected from photon shot noise. fu, fluorescence unit.

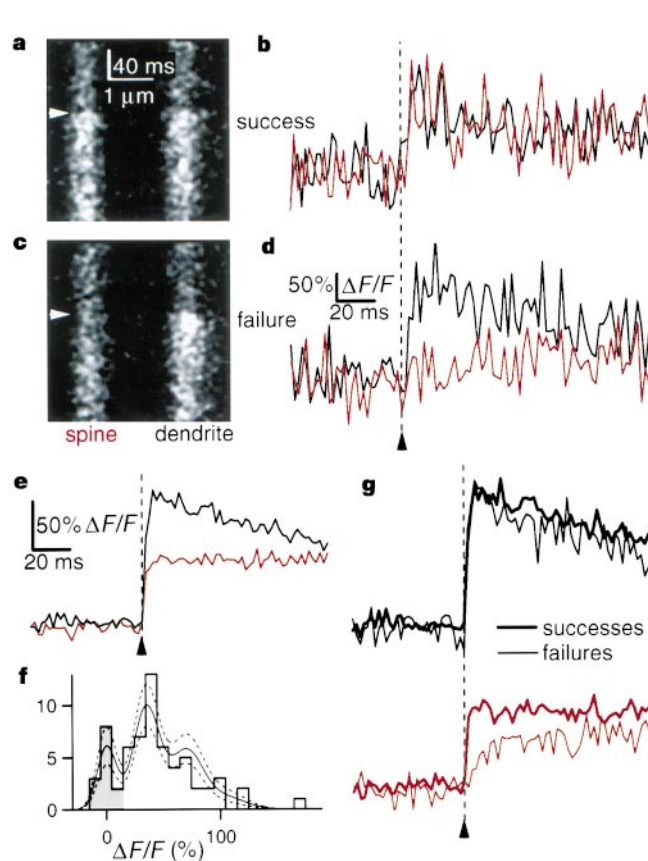


Figure 3 Failures of action-potential-evoked Ca^{2+} influx in spines. **a**, **b**, Trial showing a rapid action-potential-evoked Ca^{2+} increase in the spine and dendrite. **c**, **d**, Trial in which there was no rapid component to the Ca^{2+} increase in the spine. **e**, Average fluorescence transient (53 trials) in the spine (red) and dendrite (black). **f**, Histogram of fluorescence transient amplitudes in the spine including failures (shaded bins). Smooth lines, the mean (solid) and range (dashed, ± 1 s.d.) of the distribution predicted by a binomial model ($p = 0.43$, $N = 3$) and Poisson noise. **g**, Average fluorescence transient in the dendrite (black) and spine (red) for trials in which the spine did (thick trace, $n = 40$) or did not (thin trace, $n = 13$) show rapid Ca^{2+} influx.

mostly through R-type channels in spines and dendrites, although contributions from T-type channels cannot be excluded. This difference may reflect the smaller-calibre dendrites (<1 versus >2–3 μm in diameter) examined in our study, as well as the use of action potentials instead of step depolarizations to probe VSCC opening. In spines of neocortical pyramidal cells, most Ca^{2+} influx has been reported to be through N/P/Q-type channels¹⁵.

GABA_B receptor activation inhibits R-type VSCCs in presynaptic terminals¹⁶. G-protein-dependent modulation of VSCCs could occur through reductions in N , p (ref. 17), or i_{Ca} (ref. 18). We find that application of baclofen lowered Ca^{2+} influx in apical spines by reducing p . Baclofen increased $\text{CV}_{F,\Delta[\text{Ca}]}$ by the amount expected for a pure reduction in p ($160 \pm 8\%$ observed versus 155% expected) and by significantly more than expected for reductions in N (129%) or i_{Ca} (no increase). Thus GABA_B receptor activation probably reduces only p . Conversely, the changes in $\text{CV}_{F,\Delta[\text{Ca}]}$ and $\Delta F/F$ were used to calculate p by a method that does not rely on the failure rate or the value of k (see Methods), yielding $p_{\text{control}} = 0.59 \pm 0.03$ (range 0.50–0.66) and $p_{\text{baclofen}} = 0.33 \pm 0.04$ (range 0.23–0.40) ($n = 5$), in excellent agreement with the values obtained by analysis of failure rates. The decrease in p mediated by GABA_B can be explained either by a 50% reduction in opening probability of each VSCC or by a shift of half of the channels into a 'reluctant' state from which opening is effectively impossible following an action potential. The latter case, which is the favoured model of VSCC modulation by G proteins¹⁹, is indistinguishable by fluctuation analysis from the former when, as is the case here, the interstimulus interval is longer than the mean lifetime in the 'reluctant' state²⁰.

The confinement of GABA_B modulation of $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ to apical spines indicates that the modulatory potential of VSCCs is regulated across the expanse of the neuron (basal versus apical spines) as well as on length scales of less than 1 μm (spine head versus dendrite). This submicrometre specialization may reflect the presence of

GABA_B receptors on the spine coupled to a modulatory mechanism that is restricted to the spine head. Possible mechanisms include direct binding of membrane-delimited G-protein $\beta\gamma$ subunits to VSCCs¹⁹ or confinement of a diffusible second messenger within the spine head by the spine neck. Alternatively, the differential localization of R-type VSCC splice variants or β -subunits may determine the effects of neuromodulators on $\Delta[\text{Ca}^{2+}]$ in the dendrite and spine²¹. It has been proposed that GABA_B receptors on hippocampal pyramidal cells are primarily activated by GABA that has diffused to sites far from the inhibitory synapse²². Our results show that metabotropic effects of GABA can still be highly localized because of the inhomogeneous distribution of modulatory components.

The number of VSCCs that open in response to an action potential increases with spine head volume (Fig. 5). Assuming a roughly spherical spine head, we estimate that VSCC density in

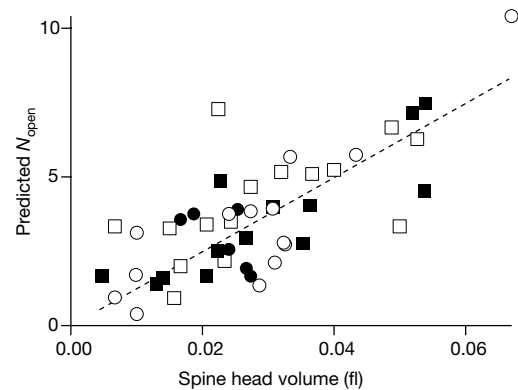


Figure 5 Number of VSCCs opened by an action potential in apical (squares) and basal (circles) spines plotted as a function of spine volume. Filled symbols are from spines in slices that were incubated in nimodipine and CTx throughout the experiment.

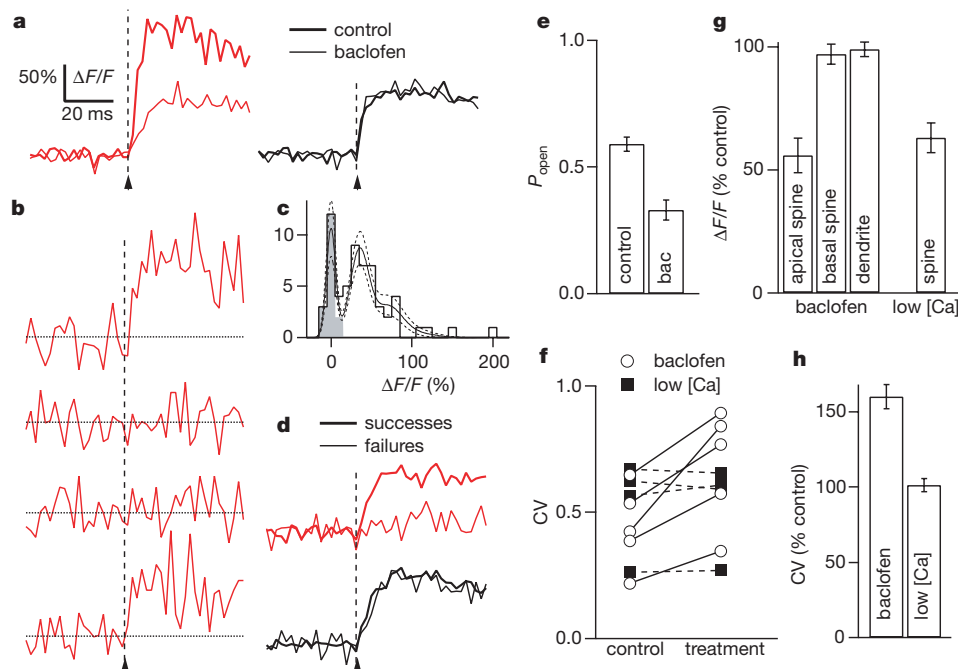


Figure 4 GABA_B receptor activation reduces Ca^{2+} influx in apical spines. **a**, Average fluorescence transient in a spine (red) and its parent dendrite (black) in control conditions ($n = 50$) and after application of baclofen ($n = 61$). **b**, Fluorescence transients in the spine in the presence of baclofen showing successes and failures (second and third traces) of action-potential-evoked Ca^{2+} influx. **c**, Histogram of fluorescence transient amplitudes in the spine including failures (shaded bins). Smooth lines, mean (solid) and range (dashed),

± 1 s.d.) of the expected distribution ($N = 3$, $p = 0.32$). **d**, Average fluorescence transient in the spine (red) and the dendrite (black) in the presence of baclofen for trials in which the spine showed a failure ($n = 17$) or success of Ca^{2+} influx ($n = 44$). **e**, Effects of baclofen on p . **f**, $\text{CV}_{F,\Delta[\text{Ca}]}$ measured from apical spines in control conditions or after the application of either baclofen (open circles) or low $[\text{Ca}^{2+}]$ ACSF (filled squares). **g**, **h**, Effects of baclofen and low $[\text{Ca}^{2+}]$ ACSF application on $\Delta F/F$ and $\text{CV}_{F,\Delta[\text{Ca}]}$.

spine heads of volume 0.05–0.10 fl is ~ 20 channels μm^{-2} , several times larger than reported for apical dendrites of these cells¹⁴. The smallest spines in our study contain less than five VSCCs and have large trial-to-trial fluctuations ($\pm 50\%$) in $\Delta[\text{Ca}^{2+}]_{\text{AP}}$. Hence stochastic channel opening influences the activation of Ca^{2+} -dependent processes and may produce noise in the induction of Ca^{2+} -dependent forms of synaptic plasticity²³.

The clear separation of failures and successes of action-potential-evoked Ca^{2+} influx (Figs 3 and 4) demonstrates that Ca^{2+} influx through a single VSCC functioning in intact neural networks can be detected. Based on the amplitude of the first peak in the amplitude histograms (Figs 3f and 4c), the opening of a single VSCC during an action potential caused a 30–40% increase in fluorescence in the spine head. Consistent with these results, a VSCC with 15 pS conductance²⁴ that opens for 200 μs is expected to increase fluorescence in a 0.1-fl spine by $\sim 45\%$, assuming typical values for endogenous Ca^{2+} buffer capacity (100), resting Ca^{2+} (60 nM), Ca^{2+} affinity for Fluo-4 (300 nM), and driving force for Ca^{2+} entry (100 mV)²⁵. Thus, coupled with fluctuation analysis, optical imaging of Ca^{2+} transients allows the study of single-channel properties in small cellular structures inaccessible to standard electrophysiological techniques. $\square$

Methods

Electrophysiology

Rat (postnatal day 14–20) hippocampal slices were cut as described²⁶. Experiments were performed at 34 °C in physiological ACSF (in mM) (127 NaCl, 25 NaHCO₃, 25 D-glucose, 2.5 KCl, 2 CaCl₂, 1 MgCl₂, 1.25 NaH₂PO₄). In some experiments one or more of the following drugs were added to the ACSF (in μM): nimodipine (10), ω -conotoxin-MVIIIC (5), baclofen (30), carboxypiperazin-propylphosphonic acid (CPP) (10). For low Ca^{2+} experiments, the ACSF contained 1 mM CaCl₂ and 2 mM MgCl₂. Membrane voltage was monitored in whole-cell current clamp. Electrodes (3–5 M Ω) contained (in mM) 135 potassium methylsulphonate, 10 HEPES, 10 sodium phosphocreatine, 4 MgCl₂, 4 Na-ATP, 0.4 Na-GTP, and 0.1–0.2 Fluo-4 (Molecular Probes). Back-propagating action potentials were induced by somatic current injection (1–3 nA, 2 ms). Data acquisition and analysis was performed using custom software written in Igor Pro (Wavemetrics).

Imaging

After cells were fully loaded with Fluo-4 (10–30 min after break-in)²⁵, individual spines located 25–120 μm (average 75 μm) from the soma on basal or apical dendrites were identified for measurement. Fluorescence from spines and their parent dendrites was measured using a custom 2PLSM microscope (software: Lucent Technologies) by scanning repeatedly (500 Hz) in a line that intersected both structures (Fig. 1). Fluorescence collected from each structure was averaged to produce fluorescence time courses for the spine head and dendrite. No smoothing or low-pass filtering was used. Summary data is given as mean $\pm$ s.e.m.; n is the number of spines. Statistical significance determined with the Student- t test at a $P < 0.05$ level.

Cells in which the surrounding slice showed detectable background fluorescence after loading were rejected. No background correction was performed other than subtraction of the photomultiplier tube dark current. This offset was measured in each trial before shutter opening (Fig. 2b). Baseline fluorescence (F_0) was determined by averaging the fluorescence in the 40 ms preceding the action potential. The shutter remained open for a further 50–100 ms after the stimulation. Peak fluorescence was measured in the 6–10 ms following the peak of the fluorescence transient. The fluorescence (F_{max}) that would arise from saturating $\Delta[\text{Ca}^{2+}]$ was estimated from the fluorescence transients evoked by trains of 30 and 40 action potentials²⁵. Relative fluorescence changes were computed as $(\Delta F/F_0)(t) = (F(t) - F_0)/F_0$.

Fluorescence measured from structures that are smaller than the two-photon excitation volume, such as spines, is proportional to the volume of the structure (assuming uniform dye filling). Spine volumes were estimated as the volume of the excitation kernel scaled by the ratio of the baseline fluorescence measured in the spine to that measured in a thick portion of the apical dendrite. The excitation volume was measured by imaging 0.1- μm -diameter fluorescent beads as ~ 0.33 fl (not shown).

Only cells with low resting $[\text{Ca}^{2+}]$ ($F_{\text{max}}/F_0 > 10$, implying $[\text{Ca}]_0 < K_D/9$) were analysed²⁵. Because peak fluorescence reached in spines and dendrites following a single action potential was $\sim 2F_0$, dye saturation is negligible and ΔF is linearly related to $\Delta[\text{Ca}^{2+}]$ to within 10–15%. $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ in CA1 spines and dendrites is through VSCCs⁴, with no contribution from calcium-induced calcium release^{26–28}.

Variance analysis

The average fluorescence transient, $\langle F(t) \rangle$, and its variance, $\sigma_F^2(t)$ were calculated from the unnormalized fluorescence from 50–200 trials (0.2 Hz) for each spine/dendrite pair. The fluctuations in the fluorescence result from three independent stochastic processes such that $\sigma_F^2 = \sigma_d^2 + \sigma_{\text{an}}^2 + \sigma_{F,\Delta[\text{Ca}]}^2$, where the component variances are due to dark noise (σ_d^2),

shot noise in the number of photons detected (σ_{an}^2), and fluctuations in $\Delta[\text{Ca}^{2+}]$ ($\sigma_{F,\Delta[\text{Ca}]}^2$). σ_d^2 was measured with the shutter closed. Variance due to shot noise obeys Poisson statistics (that is, the mean and variance of the number of photons collected are equal)²⁹. Measurements in cuvettes filled with Ca^{2+} indicators ($\sigma_{F,\Delta[\text{Ca}]}^2 = 0$) confirmed that $\sigma_F^2 = q\langle F \rangle + \sigma_d^2$, where q is the signal produced by the detection of a single photon (Fig. 2b, c). For measurements from spines and dendrites, the variance in fluorescence across trials was calculated at each of the 20 time points in the baseline period. We excluded from analysis spines in which the average of these variances was more than 1 s.e. (less than 5% of the mean variance) removed from the variance expected from Poisson and dark noise. In the remaining spines, $\sigma_{F,\Delta[\text{Ca}]}^2$ was calculated as $\sigma_{F,\Delta[\text{Ca}]}^2(t) = \sigma_F^2(t) - \sigma_d^2 - q\langle F(t) \rangle$. As ΔF and $\Delta[\text{Ca}^{2+}]$ are linearly related, the coefficient of variation of $\Delta[\text{Ca}^{2+}]$ is given by:

$$\text{CV}_{\Delta[\text{Ca}]} = \text{CV}_{F,\Delta[\text{Ca}]} = \frac{\sigma_{F,\Delta[\text{Ca}]}(t)}{\langle F \rangle - F_0} = \frac{\sqrt{\sigma_F^2 - \sigma_d^2 - q\langle F \rangle}}{\langle F \rangle - F_0}$$

Fluctuations in the number of VSCCs that open following an action potential contribute to $\sigma_{F,\Delta[\text{Ca}]}^2$. The number of open channels has mean $\langle N_{\text{open}} \rangle = Np$, variance $\sigma_{\text{open}}^2 = Np(1-p)$, and $\text{CV}_{\text{open}} = \sqrt{(1-p)/Np}$. The trial-to-trial variability in $\Delta[\text{Ca}^{2+}]$ also includes a component due to fluctuations in the number of Ca^{2+} ions that pass through an open VSCC. This contribution is analogous to multiplicative noise in photocurrent measurements³⁰. In general, $\text{CV}_{\text{open}}^2 = k\text{CV}_{F,\Delta[\text{Ca}]}^2$ where $k = (1 + \text{CV}_{F,\Delta[\text{Ca}]}^2)^{-1}$ and $\text{CV}_{F,\Delta[\text{Ca}]}$ describes the fluctuations in Ca^{2+} influx per open channel (i_{Ca}). k is a constant that represents what fraction of $\sigma_{\Delta[\text{Ca}]}^2$ is due to σ_{open}^2 and is in the range $0.5 < k < 1$: the lower bound is derived from assuming that each channel only flickers open once, with an open time that is exponentially distributed, and the upper bound from assuming that the individual channel current shows no variance. A tighter estimate of k can be derived by taking into account that VSCCs flicker open approximately four times per action potential¹⁴, implying $k > 0.8$. Thus the equation $Np = (1-p)/\text{CV}_{F,\Delta[\text{Ca}]}^2$ will underestimate the true number of open channels by at most 20%.

The probability that an action potential fails to open any VSCC in a spine is $P(0) = (1-p)^N$ and can be used, along with CV_{open} , to solve for N and p

$$P(0) = \left(\frac{\text{CV}_{\text{open}}^2}{1/n + \text{CV}_{\text{open}}^2} \right)^N \quad \text{and} \quad p = \frac{1}{1 + N\text{CV}_{\text{open}}^2}$$

Application of baclofen reduced $\Delta[\text{Ca}^{2+}]_{\text{AP}}$ in apical spines (Fig. 4). To determine if changes in i_{Ca} , N , or p cause this reduction we define the quantities:

$$R_{\text{CV}} = \frac{\text{CV}_{F,\Delta[\text{Ca}]}^2(\text{control})}{\text{CV}_{F,\Delta[\text{Ca}]}^2(\text{baclofen})} \quad \text{and} \quad R_F = \frac{(\Delta F/F)_{\text{control}}}{(\Delta F/F)_{\text{baclofen}}}$$

Changes in i_{Ca} imply $R_{\text{CV}} = 1$. Reducing N decreases R_F and increases R_{CV} such that $R_F R_{\text{CV}} = 1$. Reducing p increases R_{CV} more than it decreases R_F such that $R_F R_{\text{CV}} = (1-p_{\text{control}})(1-p_{\text{baclofen}})^{-1} < 1$. This is consistent with our experiments: we found $R_F R_{\text{CV}} = 0.67 \pm 0.07$, which matches the expected value of 0.70 calculated from p_{control} and p_{baclofen} . As a further consistency check, R_F and R_{CV} measured across control and baclofen conditions can be used to solve for p_{control} and p_{baclofen} without the need to count failures such that:

$$p_{\text{control}} = \frac{R_{\text{CV}} R_F - 1}{R_{\text{CV}} - 1}, \quad p_{\text{baclofen}} = p_{\text{control}} R_F^{-1}$$

The value of p depends on the ratio of $\text{CV}_{\Delta[\text{Ca}]}$ in the two conditions, and is thus insensitive to the value of k .

Received 19 June; accepted 25 September 2000.

- Harris, K. M. & Kater, S. B. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
- Zucker, R. S. Calcium- and activity-dependent synaptic plasticity. *Curr. Opin. Neurobiol.* **9**, 305–313 (1999).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
- Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375**, 682–684 (1995).
- Jaffe, D. B. *et al.* The spread of Na⁺ spikes determines the pattern of dendritic Ca²⁺ entry into hippocampal neurons. *Nature* **357**, 244–246 (1992).
- Svoboda, K., Tank, D. W. & Denk, W. Direct measurement of coupling between dendritic spines and shafts. *Science* **272**, 716–719 (1996).
- Hillyard, D. R. *et al.* A new Conus peptide ligand for mammalian presynaptic Ca²⁺ channels. *Neuron* **9**, 69–77 (1992).
- Triggle, D. J. & Rampe, D. 1,4-Dihydropyridine activators and antagonists: structural and functional distinctions. *Trends Pharmacol. Sci.* **10**, 507–511 (1989).
- Kavalali, E. T., Zhuo, M., Bito, H. & Tsien, R. W. Dendritic Ca²⁺ channels characterized by recordings from isolated hippocampal dendritic segments. *Neuron* **18**, 651–663 (1997).
- Barak, L. S. & Webb, W. W. Diffusion of low density lipoprotein-receptor complex on human fibroblasts. *J. Cell Biol.* **95**, 846–852 (1982).
- Katz, B. & Miledi, R. Membrane noise produced by acetylcholine. *Nature* **226**, 962–963 (1970).
- Koch, C. & Zador, A. The function of dendritic spines: devices subserving biochemical rather than electrical compartmentalization. *J. Neurosci.* **13**, 413–422 (1993).
- Christie, B. R., Eliot, L. S., Ito, K. I., Miyakawa, H. & Johnston, D. Different Ca²⁺ channels in soma and dendrites of hippocampal pyramidal neurons mediate spike-induced Ca²⁺ influx. *J. Neurophysiol.* **73**, 2553–2557 (1995).
- Magee, J. C. & Johnston, D. Characterization of single voltage-gated Na⁺ and Ca²⁺ channels in apical dendrites of rat CA1 pyramidal neurons. *J. Physiol.* **487**, 67–90 (1995).

15. Schiller, J., Schiller, Y. & Clapham, D. E. Amplification of calcium influx into dendritic spines during associative pre- and postsynaptic activation: the role of direct calcium influx through the NMDA receptor. *Nature Neurosci.* **1**, 114–118 (1998).
16. Wu, L. G., Borst, J. G. & Sakmann, B. R-type Ca^{2+} currents evoke transmitter release at a rat central synapse. *Proc. Natl Acad. Sci. USA* **95**, 4720–4725 (1998).
17. Bean, B. P. Neurotransmitter inhibition of neuronal calcium currents by changes in channel voltage dependence. *Nature* **340**, 153–156 (1989).
18. Kuo, C. C. & Bean, B. P. G-protein modulation of ion permeation through N-type calcium channels. *Nature* **365**, 258–262 (1993).
19. Zamponi, G. W. & Snutch, T. P. Modulation of voltage-dependent calcium channels by G proteins. *Curr. Opin. Neurobiol.* **8**, 351–356 (1998).
20. Boland, L. M. & Bean, B. P. Modulation of N-type calcium channels in bullfrog sympathetic neurons by luteinizing hormone-releasing hormone: kinetics and voltage dependence. *J. Neurosci.* **13**, 516–533 (1993).
21. Page, K. M., Canti, C., Stephens, G. J., Berrow, N. S. & Dolphin, A. C. Identification of the amino terminus of neuronal Ca^{2+} channel $\alpha 1$ subunits $\alpha 1B$ and $\alpha 1E$ as an essential determinant of G-protein modulation. *J. Neurosci.* **18**, 4815–4824 (1998).
22. Isaacson, J. S., Solis, J. M. & Nicoll, R. A. Local and diffuse synaptic actions of GABA in the hippocampus. *Neuron* **10**, 165–175 (1993).
23. Petersen, C. C., Malenka, R. C., Nicoll, R. A. & Hopfield, J. J. All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl Acad. Sci. USA* **95**, 4732–4737 (1998).
24. Tottene, A., Moretti, A. & Pietrobon, D. Functional diversity of P-type and R-type calcium channels in rat cerebellar neurons. *J. Neurosci.* **16**, 6353–6363 (1996).
25. Maravall, M., Mainen, Z. M., Sabatini, B. & Svoboda, K. Estimating intracellular calcium concentrations and buffering without wavelength ratioing. *Biophys. J.* **78**, 2655–2667 (2000).
26. Mainen, Z. F., Malinow, R. & Svoboda, K. Synaptic calcium transients in single spines indicate that NMDA receptors are not saturated. *Nature* **399**, 151–155 (1999).
27. Markram, H., Helm, P. J. & Sakmann, B. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. *J. Physiol.* **485**, 1–20 (1995).
28. Emptage, N., Bliss, T. V. P. & Fine, A. Single synaptic events evoke NMDA receptor-mediated release of calcium from internal stores in hippocampal dendritic spines. *Neuron* **22**, 115–124 (1999).
29. Rice, S. O. in *Noise and Stochastic Processes* (ed. Wax, N.) 133–294 (Dover, New York, 1954).
30. Jones, R., Oliver, C. J. & Pike, E. R. Experimental and theoretical comparison of photon-counting and current measurements of light intensity. *Appl. Opt.* **10**, 1673–1680 (1970).

Acknowledgements

We thank B. J. Burbach and P. O'Brien for help with experiments, and Z. Mainen, R. Malinow, M. Maravall, W. Regehr, R. Weinberg and R. Yasuda for comments on the manuscript. This work was supported by the Pew, Klingenstein and Mathers Foundations, the Howard Hughes Medical Institute, NIH and a Helen Hay Whitney Fellowship.

Correspondence and requests for materials should be addressed to K.S. (e-mail: svoboda@cshl.org).

Role of cortical tumour-suppressor proteins in asymmetric division of *Drosophila* neuroblast

Tomokazu Ohshiro^{*†}, Takako Yagami^{*}, Chuan Zhang^{*†} & Fumio Matsuzaki^{*†‡}

^{*}CREST, Japan Science and Technology Corporation.

[‡]Cell Asymmetry Group, RIKEN Center for Developmental Biology.

[†]Department of Developmental Neurobiology, Institute of Development, Aging, and Cancer, Tohoku University, 4-1 Seiryomachi, Aoba-ku, Sendai 980-8575, Japan

Cellular diversity during development arises in part from asymmetric divisions, which generate two distinct cells by transmitting localized determinants from a progenitor cell into one daughter cell. In *Drosophila*, neuroblasts undergo typical asymmetric divisions to produce another neuroblast and a ganglion mother cell^{1,2}. At mitosis, neural fate determinants, including Prospero and Numb, localize to the basal cortex^{3,4}, from which the ganglion mother cell buds off; Inscuteable and Bazooka, which regulate spindle orientation, localize apically^{5–8}. Here we show that a tumour-suppressor protein, Lethal giant larvae (Lgl)⁹, is essential for asymmetric cortical localization of all basal determinants in mitotic neuroblasts, and is therefore indispensable for neural fate

decisions. Lgl, which itself is uniformly cortical, interacts with several types of Myosin to localize the determinants. Another tumour-suppressor protein, Lethal discs large (Dlg)¹⁰, participates in this process by regulating the localization of Lgl. The localization of the apical components is unaffected in *lgl* or *dlg* mutants. Thus, Lgl and Dlg act in a common process that differentially mediates cortical protein targeting in mitotic neuroblasts, and that creates intrinsic differences between daughter cells.

In mitotic neuroblasts, the Prospero transcription factor and Numb, an antagonist of Notch signalling, associate with their respective adapter proteins, Miranda^{11,12} and Partner of Numb (Pon)¹³, and thereby localize to the basal cortex. In contrast, Inscuteable (Insc), Bazooka (Baz) and Partner of Inscuteable (Pins) form a ternary complex at the apical cortex independently of the basal determinants^{5–8}. However, the mechanisms that underlie the asymmetric protein sorting in neuroblasts are not known. To address this issue, we have searched for chromosomal deficiencies that affect the subcellular distribution of Miranda. Such screening identified the *lgl* tumour-suppressor gene^{9,14,15}, which encodes a protein containing WD40 repeats. In wild-type neuroblasts, Miranda, which localizes apically during interphase, accumulates at the basal cortex upon mitosis after a transient spread into the cytoplasm^{11,12} (Fig. 1a). In germline clone embryos lacking both maternal and zygotic *lgl* activity (*lgl*^{GLC} embryos), Miranda did not localize asymmetrically in mitotic neuroblasts, but rather was distributed uniformly throughout the cortex as well as in the cytoplasm, where it was concentrated along microtubule structures (Fig. 1b, e, f). Consequently, Miranda segregated into both the

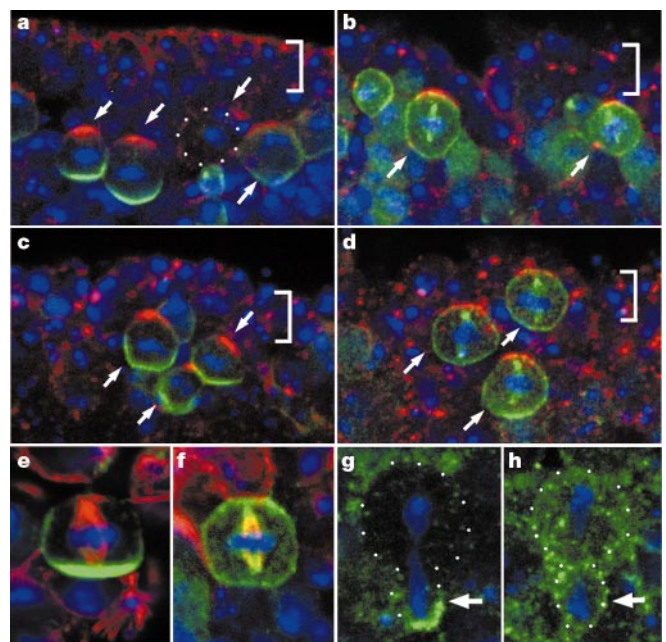


Figure 1 Miranda and Numb but not Baz mislocalize in mitotic neuroblasts of *lgl* and *dlg* mutant embryos. Apical is up and DNA is shown in blue in all figures. **a–d**, Miranda (green) and Baz (red) in wild-type embryos (**a**), *lgl*^{GLC} embryos (**b**), *lgl*^{GLC} embryos that specifically express Lgl in neuroblasts (**c**) and *dlg*^{GLC} embryos (**d**) at stage 11. Brackets mark the epithelial layer; arrows indicate mitotic neuroblasts. We note that in epithelial cells Baz mislocalized as discrete cortical patches in *lgl*^{GLC} (**b**) and *dlg*^{GLC} (**d**) embryos, in contrast to its localization at the apical margin of lateral cortex and the apical surface in wild-type epithelia (**a**). Expression of the *lgl*/transgene in neuroblasts restored the normal Miranda crescent leaving Baz mislocalized in epithelial cells (**c**). **e, f**, Miranda (green) and α -tubulin (red) in wild-type (**e**) and *lgl*^{GLC} (**f**) neuroblasts at metaphase. The colocalization of Miranda with the mitotic spindle depends on the microtubule as revealed by the treatment of a microtubule-depolymerizing drug, Colcemid (data not shown). **g, h**, Numb (green) in wild-type (**g**) and *lgl*^{GLC} (**h**) neuroblasts at telophase. Arrows indicate budding ganglion mother cells (GMCs).

daughter neuroblast and the ganglion mother cell (GMC). Numb (Fig. 1g, h) and Pon (not shown) were also distributed uniformly at the cortex and in the cytoplasm.

Lgl is involved in epithelial polarity^{9,16}, and neuroblasts inherit Baz as an apical polarity cue during the formation from neuroepithelia^{6,7}. In epithelia lacking *lgl* function, Baz mislocalized as irregular patches (Fig. 1b). Baz as well as Insc and Pins normally localized as an apical crescent in the mutant neuroblasts (Fig. 1b; and data not shown), however, suggesting that the apical cue is inherited to neuroblasts in the absence of *lgl*. Moreover, neuroblast-specific expression of the *lgl* transgene could restore normal localization of Miranda (Fig. 1c; and see Methods). Thus, Lgl probably functions autonomously within neuroblasts for Miranda localization, and the phenotype in *lgl* neuroblasts would not result from the inheritance of an abnormal polarity cue. Together, we conclude that Lgl acts in the localization of the basal determinants in neuroblasts, but not of the apical Baz–Insc–Pins complex.

We next investigated whether other tumour-suppressor genes contribute to protein localization in neuroblasts. The tumour-suppressor gene *dlg* encodes a membrane-associated guanylate kinase homologue¹⁰. Germline clone embryos lacking both maternal and zygotic *dlg* activity (*dlg*^{GLC} embryos) exhibited defective localization of Miranda (Fig. 1d) and Numb (data not shown) essentially identical to that of *lgl*^{GLC} embryos, suggesting that both tumour-suppressor proteins function in the same process in neuroblasts. To investigate the relationship between the roles of Dlg and Lgl, we compared their subcellular localization in neuroblasts. Both Lgl and Dlg were distributed mainly throughout the cortex, whereas the amount of Lgl in the cytoplasm appeared to be greater in mitosis than in interphase (Fig. 2a, b). The cortical localization of Lgl appears to be important for its function—whereas the mutant protein encoded by the temperature-sensitive allele *lgl*^{ts3} (ref. 16) was distributed normally at the permissive temperature (18 °C), it failed to localize cortically at the restrictive temperature (29 °C) (Fig. 2c; and ref. 17). The wild-type Lgl protein exhibited a similar, abnormal cytoplasmic distribution in *dlg*^{GLC} embryos (Fig. 2d), whereas Dlg localization was not affected in *lgl*^{GLC} embryos (data not shown). Thus, the cortical localization of Lgl requires *dlg* activity, suggesting that Dlg may function in localization of cell-fate determinants in neuroblasts by positioning Lgl at the cortex.

To distinguish whether Lgl establishes or maintains determinant localization, we performed temperature-shift experiments with the *lgl*^{ts3} allele. At 18 °C, most (97%, *n* = 262) *lgl*^{ts3}/*lgl*^{ts3} embryos of females homozygous for *lgl*^{ts3} exhibited normal Miranda localization at the basal cortex during metaphase and telophase (Fig. 3a).

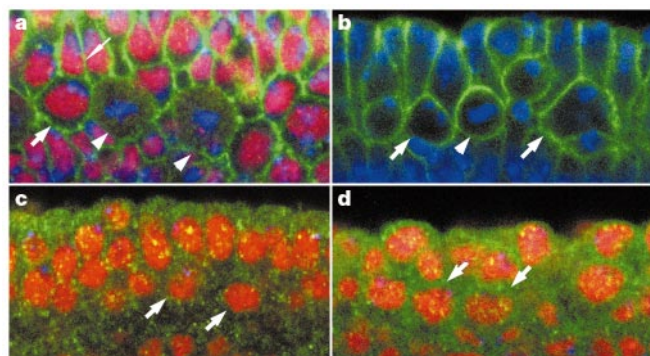


Figure 2 Localization of Lgl and Dlg in neuroblasts. Wild-type embryos (a, b), *lgl*^{ts3} embryos at 29 °C (c) and *dlg*^{GLC} embryos (d) were stained with the anti-Lgl-N antibody (green) in a, c and d, and with the anti-Dlg antibody (green) in b. Anti-Cen 190 staining (red) in a, c and d reveals nuclei at interphase and centrosomes at mitosis. The *lgl*^{ts3} mutant protein at the restrictive temperature (c) and the wild-type Lgl protein in *dlg*^{GLC} embryos (d) show a cytoplasmic distribution. Arrowheads indicate metaphase neuroblasts; thick arrows, interphase neuroblasts; thin arrow, epithelial cell.

Ten minutes after shifting to 29 °C, mislocalization of Miranda was apparent in 61% of metaphase or anaphase neuroblasts (*n* = 159) as well as in 43% of telophase neuroblasts (*n* = 124), whereas the apical localization of Miranda in interphase was not affected (Fig. 3b, c). On the basis of live recordings of basal crescent formation in cells that express a fusion protein comprising Miranda and green fluorescent protein (see Methods), we estimated that the time required for neuroblast mitosis is 14.3 ± 2.4 min (mean $\pm$ s.d., *n* = 13). Thus, a temperature shift during mitosis was able to induce mislocalization of Miranda, as apparent in telophase neuroblasts, indicating that Lgl functions during mitosis to determine Miranda localization. Metaphase arrest can be induced in *lgl*^{ts3}/*lgl*^{ts3} embryos by introduction of the *fizzy* (*fzy*) mutation¹⁸. The localization of Miranda in such metaphase-arrested neuroblasts was virtually insensitive to the shift to 29 °C (Fig. 3d). Thus, Miranda is not affected by the decrease in *lgl* function after it has localized to the basal cortex. These results therefore indicate that Lgl may act early during mitosis to recruit Miranda to the cortex but does not contribute to the maintenance of Miranda in this location.

Because Lgl is a component of cortical protein complexes that include nonmuscle Myosin II, or Zipper (Zip)¹⁹, we tested genetic interactions between *lgl* and *zip* in Miranda localization by examining embryos zygotically mutant for both *lgl* and *zip* (*lgl*^{ts3}/*lgl*^{ts3}, *zip*¹/*zip*¹ (*lgl*–*zip*) embryos)²⁰. The *zip*¹ mutation did not affect Miranda localization throughout embryonic development and *lgl*–*zip* embryos showed no difference in Miranda localization from zygotic *lgl*^{ts3} embryos until late embryonic stages (stage 16) owing to maternal contribution of *zip*. At stage 17 when maternal *zip* had been exhausted, however, *lgl*–*zip* embryos appeared to restore the basal crescent of Miranda in metaphase neuroblasts (62%, *n* = 150), whereas zygotic *lgl*^{ts3} embryos at the same stage did not (2%, *n* = 138) (Fig. 3e, f). Thus, Lgl might act for Miranda localization in part by suppressing *zip* function directly or indirectly, consistent with a study on yeast²¹ that indicated negative genetic interactions between

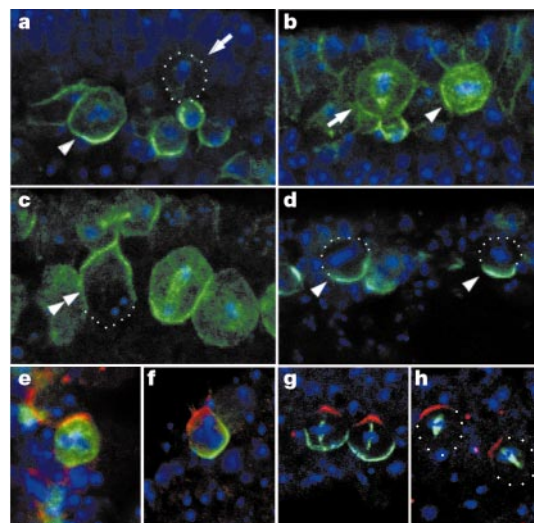


Figure 3 Requirement for Lgl early during mitosis and interactions with Myosins in Miranda localization. a–d, Miranda (green) distribution in *lgl*^{ts3}/*lgl*^{ts3} embryos of homozygous *lgl*^{ts3} females at 18 °C (a) or 10 min after a shift to 29 °C (b, c), and in *lgl*^{ts3}/*lgl*^{ts3}, *fzy*^{+/+}/*fzy*^{+/+} embryos of *lgl*^{ts3}/*lgl*^{ts3}, *fzy*^{+/+}/*fzy*^{+/+} females, which were shifted to 29 °C for 10 min at stage 14 (d). Arrowheads indicate metaphase neuroblasts; arrows, telophase neuroblasts; double arrowhead, interphase neuroblast. e, f, Miranda (green) and Insc (red) distribution in mitotic neuroblasts of zygotic *lgl*^{ts3}/*lgl*^{ts3} (e) and *lgl*^{ts3}/*lgl*^{ts3}, *zip*¹/*zip*¹ (f) embryos at the late embryonic stage (stage 17). g, h, Distribution of Miranda (green) and Insc (red) in neuroblasts of wild-type (g) and *lgl*^{GLC} (h) embryos, both of which were treated with BDM (50 mM). Note that Insc remained apical in *lgl*^{GLC} embryos even after BDM treatment, indicating that apical protein localization is independent of *lgl* and myosin function.

Lgl homologues and Myosin II. Alternatively, the asymmetric distribution of Pon requires *myosin* function in neuroblasts²², as revealed by the use of 2,3-butanedione monoxime (BDM) that generally inhibits *myosin* function²³. We also examined the effect of BDM on Miranda localization. Treatment of wild-type embryos with BDM phenocopied *lgl* mutants, resulting in a partial redistribution of Miranda from the cortex to microtubules (Fig. 3g). The effect of BDM was more marked in *lgl^{GLC}* embryos: as the BDM concentration increased, the relocalization of Miranda to microtubules was synergistically enhanced in most BDM-treated neuroblasts and resulted in the complete exclusion of Miranda from the cortex at 50 mM BDM (Fig. 3h). The phenocopy and enhancement of *lgl* mutations by general inhibition of *myosin* function are in contrast with the suppressive effects of *zip* mutations, suggesting that Lgl cooperates with at least one type of Myosin other than Zip to anchor Miranda at the cell cortex. We thus infer that Lgl regulates negatively *myosin II* function and also positively the function of another Myosin isotype in cortical protein targeting in neuroblasts.

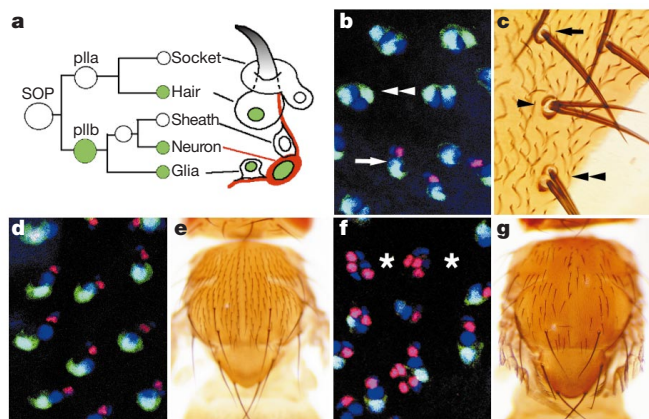


Figure 4 Cell-fate transformation in the external sensory organ lineage induced by a reduction in *lgl* activity. **a**, Cell lineage of the external sensory organ. At each division, Numb segregates into one daughter cell (green), in which Notch signalling is inhibited^{3,24,25}. **b, d, f**, Cell clusters of the sensory organ precursor progeny in pupal notum. The A101 enhancer trap of *neuralized* marks all sensory cells²⁸ (blue); Suppressor of Hairless (Su(H), green), socket cells²⁸, Elav (red), neurons. **c, e, g**, external sensory organs in adult notum. **b, c**, *lgl^{ts3}* animals reared at 29 °C during external sensory organ development. Occasional transformation from pIIb to pIIa produces sensory cell clusters containing two socket cells without a neuron (double arrowhead in **b**) and resulted in aberrant external sensory organs, consisting of the duplicated hair-socket pairs (arrowhead and double arrowhead in **c**). Arrows indicate normal clusters of sensory cells (**b**) and external sensory organs with a hair and a socket (**c**). In some external sensory organs, two sockets become fused (arrowhead in **c**). **d, e**, *lgl^{ts3}* animals carrying a *numb* transgene under the control of the heat shock promoter (*hs-numb*). Transient shifts to 29 °C during external sensory organ development do not perturb the formation of normal sensory cell clusters containing a neuron and a socket cell (**d**), resulting in the normal external sensory organs (**e**). In **d**, glial cells have migrated off the clusters at this stage. **f, g**, *lgl^{ts3}* animals carrying a *hs-numb* gene. The same temperature shift frequently generates sensory cell clusters, in which inner cells (for example, neurons) increase at the expense of outer cells (for example, socket) as marked by asterisks in **f**. Most adults (88%, *n* = 291) consequently show frequent loss of the external sensory structure (**g**).

Table 1 Effects of reduced *Lgl* activity on sensory cell fates

Genotype	Temperature (°C)	Number of individuals with duplicated hairs in notum (%) ^a	Average number of duplicated hairs per notum
<i>lgl^{ts3}/lgl^{ts3}, numb⁺/numb⁺</i>	18	0 (<i>n</i> = 50)	0
<i>lgl^{ts3}/lgl^{ts3}, numb⁺/numb⁺</i>	29	83 (<i>n</i> = 102)	2.9
<i>lgl^{ts3}/lgl^{ts3}, numb⁺/numb⁺</i>	18	7 (<i>n</i> = 57)	1.0
<i>lgl^{ts3}/lgl^{ts3}, numb⁺/numb⁺</i>	29	100 (<i>n</i> = 14)	32

^a Microchaetae with duplicated hairs in notum were counted.

We would expect the abnormal distribution of Numb and Miranda in *lgl* mutant neuroblasts to result in incorrect determination of neural cell fate. Given the difficulty of monitoring neural cell fate in severely distorted *lgl^{GLC}* embryos, we tested this prediction by analysing the lineage of the external sensory organ in the notum, in which all cell divisions are asymmetric and sibling cells adopt distinct fates as a result of the asymmetric inheritance of Numb^{3,24,25} (Fig. 4a). Sensory organ precursor cells in this lineage segregate Numb into a daughter cell pIIb, which subsequently generates three inner cells (a glial cell, a neuron and a sheath cell). The sibling pIIa cell divides into two outer cells constituting the external sensory structure, a hair and a socket. Exposure of *lgl^{ts3}* mutant larvae to 29 °C during external sensory organ development mislocalized Numb in mitotic precursor cells, as observed in neuroblasts (data not shown), and often transformed inner cells into outer cells resulting in duplicated external sensory structures, a phenotype expected from loss of *numb* function³ (Fig. 4b, c). Indeed, this notum phenotype was enhanced by reducing the *numb* gene dosage by half (*lgl^{ts3}/lgl^{ts3}, numb⁻/numb⁺* animals) (Table 1). Equal partition of Numb between sibling cells would result in *numb* gain of function phenotypes because the half dose of *numb* is enough for correct cell-fate decisions. The observed *numb* loss-of-function phenotype therefore suggests that a reduction in *lgl* activity does not only equalize Numb distribution between sibling cells but also attenuates *numb* function, consistent with our observation of cytoplasmic Numb in the *lgl* mutants. Conversely, the presence of an extra *numb* gene induced opposite phenotypes under the *lgl* mutant condition (Fig. 4d–g). The outer cells were frequently transformed into the inner cells, resulting in the loss of the external sensory structure (Fig. 4f, g). This appearance of the *numb* gain of function phenotypes is simply explained by the fact that the partition of additional Numb from the transgene into both sibling cells raises *numb* activities over the threshold necessary to suppress Notch function in both cells. Our data thus indicate that Lgl is essential in neural fate decisions through cortical targeting of cell-fate determinants.

There are two important processes associated with the asymmetric division^{1,2}: first, the asymmetric localization of cell-fate determinants, which is achieved by specific adapter proteins that themselves localize asymmetrically to the cortex in neuroblasts^{3,4}; and second, the orientation of the mitotic spindle and its coordination with the polarized localization of the determinants, which require the apical Baz–Insc–Pins complex^{5–8}. Our study has revealed another important process mediated by Dlg, Lgl and Myosins, which is responsible for the cortical anchoring of the determinant–adapter complexes. This process occurs upstream of the first and independently or parallel to the second of those two aspects of asymmetric division, as the localization of Lgl and Dlg is independent of apical or basal components (data not shown). Both Lgl and Dlg contribute to the generation or maintenance of epithelial polarity^{9,16,17,26,27}, and zygotic mutants of the corresponding genes develop epithelial cell tumours as well as brain tumours at late larval stages^{9,10,26,27}. These previous observations with epithelial cells, together with our data on the roles of Lgl and Dlg in protein targeting in neuroblasts, suggest that aberrant sorting of intracellular proteins may be responsible for the tumour formation apparent in larval stages of *lgl* and *dlg* mutants. □

Methods

Fly stocks and genetics

We used *lgl¹* and *lgl⁴* as null alleles for deletion of the *lgl* transcriptional unit¹⁴, *lgl^{ts3}* as a temperature-sensitive allele¹⁶, *zip¹* (a strong allele) and *zip^{IIIX62}* (a deficiency uncovering *zip* and *gooseberry*) as *zip* alleles²⁰, *numb²* as a null allele (a gift from J. Skeath), *dlg^{m52}* (ref. 10), *fzy¹* (ref. 18), the *hs-numb* transformant³, and the A101enhancer trap line²⁸. Germine clone embryos for *lgl¹* (*lgl^{GLC}* embryos) and for *dlg^{m52}* (*dlg^{GLC}* embryos) were produced by the FLP–DFS technique²⁹. Genetic interactions between *lgl* and *zip* were examined for embryos of a genotype *lgl⁴/lgl⁴, zip¹/zip^{IIIX62}*. All mutations were balanced by blue balancer chromosomes to distinguish heterozygous embryos.

Rescue experiment

The full-length *lgl* complementary DNA (LD06034 obtained from BDGP) was used to construct an *UAS-lgl* transgene. The *lgl* transgene was expressed by using the *UAS-GAL4* system³⁰ and a neuroblast-specific *GAL4* driver, *pros-GAL4* (Goto, S. and F.M., unpublished data).

BDM treatment

Dechorionated embryos were gently shaken for 30 min in a 1:1 mixture of *n*-octane and Schneider's medium (Gibco BRL) containing various concentrations of BDM (Sigma). The medium was subsequently replaced by an equal volume of 4% paraformaldehyde for fixation.

Time-lapse recording of Miranda-GFP

A *UAS-Miranda-GFP* fusion gene was constructed using a full-length *miranda* cDNA and a *GFP* gene (S65T/I167T, provided by A. Brand) and expressed in embryos produced from crosses of transformants carrying the *UAS-Miranda-GFP* gene with the *daughterless-GAL4* driver (a gift from E. Knust). Formation of the basal crescent of the fusion protein was monitored in living embryos with a confocal microscope (BioRad MRC1024).

Sensory lineage analysis

lgl^{ts3} or *lgl*^{ts3}, A101/TM6B females were separately mated with males of genotypes *lgl*^{ts3}, *lgl*^{ts3}, *numb*^{ts3}/CyO; or *lgl*^{ts3}, *hs-numb*/CyO. After laying, eggs were grown at 18 °C. Early third instar larvae were picked up and reared at 29 °C until 30 h after pupation for dissecting pupae or until eclosion for observing adult nota.

Generation of antibodies

The anti-Lgl-N antibody was raised against the amino-terminal portion (KGQPSADRHRLQKD) of Lgl in rabbits and affinity purified for immunostaining; mouse monoclonal anti-Mira-81 against the carboxy-terminal portion (SPPQKQVLKARNI) of Miranda; rabbit antibodies to Numb against the polypeptide KQLSPDLPISTAR; and rabbit anti-Baz against the C-terminal PSQYGSAGSQ-PHASKV. The specificity of all antibodies was confirmed by immunoblot analysis or by immunohistochemical staining of null mutants for the corresponding genes.

Immunohistochemistry

Embryos were fixed with 4% paraformaldehyde for 20 min or with 37% formaldehyde for 3 min (to visualize microtubules). Primary antibodies were as follows: rabbit anti-Mira-C¹¹, mouse anti-Mira-81, rabbit anti-Numb, rabbit anti-Lgl-N, rabbit anti-Baz, mouse anti- α -tubulin DM1a (Sigma), guinea pig anti-Dlg (provided by P. Bryant), mouse anti-Cen 190 Bx63 (from D. Glover), rabbit anti-Insc⁵, mouse anti-Elav 9F8A9 (Developmental Studies of Hybridoma Bank, University of Iowa) and rat anti-Su (H) (ref. 28). Chromosomes were visualized with TOTO3 (Molecular Probes). Confocal images were obtained as described¹¹.

Received 4 August; accepted 14 September 2000.

- Jan, L. Y. & Jan, Y. N. Asymmetric cell division. *Nature* **392**, 775–778 (1998).
- Matsuzaki, F. Asymmetric division of *Drosophila* neural stem cells: a basis for neural diversity. *Curr. Opin. Neurobiol.* **10**, 38–44 (2000).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of Numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- Hirata, J., Nakagoshi, H., Nabeshima, Y. & Matsuzaki, F. Asymmetric segregation of the homeodomain protein Prospero during *Drosophila* development. *Nature* **377**, 627–630 (1995).
- Kraut, R., Chia, W., Jan, L. Y., Jan, Y. N. & Knoblich, J. A. Role of *inscutable* in orienting asymmetric cell divisions in *Drosophila*. *Nature* **383**, 50–55 (1996).
- Wodarz, A., Ramrath, A., Kuchinke, U. & Knust, E. Bazooka provides an apical cue for Inscutable localization in *Drosophila* neuroblasts. *Nature* **402**, 544–547 (1999).
- Schober, M., Schaefer, M. & Knoblich, J. A. Bazooka recruits Inscutable to orient asymmetric cell divisions in *Drosophila* neuroblasts. *Nature* **402**, 548–551 (1999).
- Yu, F., Morin, X., Cai, Y., Yang, X. & Chia, W. Analysis of *partner of inscutable*, a novel player of *Drosophila* asymmetric divisions, reveals two distinct steps in Inscutable apical localization. *Cell* **100**, 399–409 (2000).
- Gateff, E. Malignant neoplasms of genetic origin in *Drosophila melanogaster*. *Science* **200**, 1448–1459 (1978).
- Woods, D. F. & Bryant, P. J. The discs-large tumor suppressor gene of *Drosophila* encodes a guanylate kinase homolog localized at septate junctions. *Cell* **66**, 451–464 (1991).
- Ishikawa-Kataoka, H., Skeath, J. B., Nabeshima, Y., Doe, C. Q. & Matsuzaki, F. Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* **390**, 625–629 (1997).
- Shen, C. P., Jan, L. Y. & Jan, Y. N. Miranda is required for the asymmetric localization of Prospero during mitosis in *Drosophila*. *Cell* **90**, 449–458 (1997).
- Lu, B., Rothenberg, M., Jan, L. Y. & Jan, Y. N. Partner of Numb colocalizes with Numb during mitosis and directs Numb asymmetric localization in *Drosophila* neural and muscle progenitors. *Cell* **90**, 225–235 (1998).
- Mechler, B. M., McGinnis, W. & Gehring, W. J. Molecular cloning of *lethal(2)giant larvae*, a recessive oncogene of *Drosophila melanogaster*. *EMBO J.* **4**, 1551–1557 (1985).
- Jacob, L., Oppen, M., Metzroth, B., Phannavong, B. & Mechler, B. M. Structure of the *l(2)gl* gene of *Drosophila* and delimitation of its tumor suppressor domain. *Cell* **50**, 215–225 (1987).
- Manfrulli, P., Arquier, N., Hanratty, W. P. & Semeriva, M. The tumor suppressor gene, *lethal(2)giant larvae* (*l(2)gl*), is required for cell shape change of epithelial cells during *Drosophila* development. *Development* **122**, 2283–2294 (1996).
- De Lorenzo, C., Strand, D. & Mechler, B. M. Requirement of *Drosophila l(2)gl* function for survival of

the germline cells and organization of the follicle cells in a columnar epithelium during oogenesis. *Int. J. Dev. Biol.* **43**, 207–217 (1999).

- Dawson, I. A., Roth, S., Akam, M. & Artavanis-Tsakonas, S. Mutations of the *fizzy* locus cause metaphase arrest in *Drosophila melanogaster* embryos. *Development* **117**, 359–376 (1993).
- Strand, D. *et al.* The *Drosophila* lethal(2)giant larvae tumor suppressor protein forms homo-oligomers and is associated with nonmuscle myosin II heavy chain. *J. Cell Biol.* **127**, 1361–1373 (1994).
- Young, P. E., Richman, A. M., Ketchum, A. S. & Kiehart, D. P. Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function. *Genes Dev.* **7**, 29–41 (1993).
- Kagami, M., Toh-e, A. & Matsui, Y. Sro7p, a *Saccharomyces cerevisiae* counterpart of the tumor suppressor *l(2)gl* protein, is related to myosins in function. *Genetics* **149**, 1717–1727 (1998).
- Lu, B., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of Numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
- Cramer, L. P. & Mitchison, T. J. Myosin is involved in postmitotic cell spreading. *J. Cell Biol.* **131**, 179–189 (1995).
- Wang, S., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. Only a subset of the binary cell fate decisions mediated by Numb/Notch signaling in *Drosophila* sensory organ lineage requires Suppressor of Hairless. *Development* **124**, 4435–4446 (1997).
- Gho, M., Bellaiche, Y. & Schweisguth, F. Revisiting the *Drosophila* microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development* **126**, 3573–3584 (1999).
- Woods, D. F., Hough, C., Peel, D., Callaini, G. & Bryant, P. J. Dlg protein is required for junction structure, cell polarity, and proliferation control in *Drosophila* epithelia. *J. Cell Biol.* **134**, 1469–1482 (1996).
- Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. *Science* **289**, 113–116 (2000).
- Gho, M., Lecourtis, M., Geraud, G., Posakony, J. W. & Schweisguth, F. Subcellular localization of Suppressor of Hairless in *Drosophila* sense organ cells during Notch signalling. *Development* **122**, 1673–1682 (1996).
- Chou, T. B. & Perrimon, N. Use of a yeast site-specific recombinase to produce female germline chimeras in *Drosophila*. *Genetics* **131**, 643–653 (1992).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).

Acknowledgements

We thank A. Brand, P. Bryant, W. Chia, D. Glover, S. Hayashi, Y.N. Jan, E. Knust, C. Peng, F. Schweisguth, M. Semeriva, A. Shearn, J. Skeath, Umea *Drosophila* Stock Center and Bloomington *Drosophila* Stock Center for providing flies and reagents; K. Hisata for technical assistance; N. Fuse, Y. Izumi, A. Sehara-Fujisawa and W. Chia for comments on the manuscript. We also thank C. Peng and C. Doe for communicating unpublished results; we are indebted to them for first observing Miranda in *lgl-zip* double mutants. This work was supported (through F.M.) by grants-in-aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan, and by CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation.

Correspondence and requests for materials should be addressed to F.M. (e-mail: fumio@idac.tohoku.ac.jp).

The tumour-suppressor genes *lgl* and *dlg* regulate basal protein targeting in *Drosophila* neuroblasts

Chian-Yu Peng, Laurina Manning, Roger Albertson & Chris Q. Doe

Institute of Molecular Biology, Institute of Neuroscience, Howard Hughes Medical Institute, University of Oregon 1254, Eugene, Oregon 97403, USA

Drosophila neuroblasts are a model system for studying asymmetric cell division: they divide unequally to produce an apical neuroblast and a basal ganglion mother cell that differ in size, mitotic activity and developmental potential. During neuroblast mitosis, an apical protein complex orients the mitotic spindle and targets determinants of cell fate to the basal cortex¹, but the mechanism of each process is unknown. Here we show that the tumour-suppressor genes *lethal giant larvae* (*lgl*) and *discs large* (*dlg*) regulate basal protein targeting, but not apical complex formation or spindle orientation, in both embryonic and larval neuroblasts. Dlg protein is apically enriched and is required for maintaining cortical localization of Lgl protein. Basal protein targeting requires microfilament and myosin function, yet the *lgl* phenotype is strongly suppressed by reducing levels of myosin II. We conclude that Dlg and Lgl promote, and

myosin II inhibits, actomyosin-dependent basal protein targeting in neuroblasts.

Embryonic *Drosophila* neuroblasts develop from an apical/basal polarized epithelium. Individual cells delaminate into the embryo, enlarge to form neuroblasts, and begin a series of asymmetric cell divisions; these divisions result in the production of a large mitotically active apical cell (neuroblast), and a smaller basal cell (ganglion mother cell, GMC) that differentiates into two neurons or glia¹. A growing number of proteins are known to be asymmetrically localized in mitotic neuroblasts: apically localized proteins include Bazooka (Baz), Inscuteable (Insc) and Partner of Inscuteable (Pins); while basally targeted proteins include Miranda, Prospero, Partner of Numb (Pon) and Numb, which are important for GMC development^{1,2}. Miranda and Prospero are apically localized at late interphase before their mitosis-dependent transport to the basal cortex³. The Baz/Insc/Pins apical complex is required for both apical/basal spindle orientation and basal protein targeting^{4–7}, but little is known about how this complex regulates either process.

To identify genes required for apical/basal protein targeting in neuroblasts, we screened deficiency stocks looking for defects in Prospero basal localization in neuroblasts. This screen identified the *lethal giant larvae* (*lgl*) gene, which encodes a WD-40 repeat protein with homologues in many species, including the closely related 'Lgl family' genes *Lgl1/Lgl2* (human), *Lgl1* (mouse), *U51993*

(*Caenorhabditis elegans*); the slightly more divergent 'Tomosyn family' genes *Tomosyn* (rat), *KIAA1006* (human), *C617762* (*Drosophila*), and *M01A10* (*C. elegans*); and recently duplicated genes similar to both families, *sro7/sro77* (budding yeast)^{8–11}. In *Drosophila*, *lgl* mutations affect protein targeting to epithelial apical junctions¹², epidermal cell-shape changes¹³, and produce tumours of the brain and the imaginal disc¹³. This spectrum of phenotypes has been noted for another tumour-suppressor gene, *discs large* (*dlg*)¹⁴. Here we explore the role of Lgl and Dlg in regulating neuroblast cell polarity.

We compared apical and basal protein targeting in neuroblasts from wild-type embryos and embryos that lack all maternal and zygotic Lgl or Dlg function (called *lgl^{GLC}* or *dlg^{GLC}* embryos; see Methods). Wild-type metaphase neuroblasts show apical Insc/Pins localization, and basal Miranda/Prospero/Pon crescents (Fig. 1a–d; Table 1). In addition, Miranda and Prospero proteins can be observed around the apical centrosome and weakly on the mitotic spindle in wild-type neuroblasts (Fig. 1b; Supplementary Information Fig. 1a–h). In contrast, all *lgl^{GLC}* and *dlg^{GLC}* metaphase neuroblasts show cytoplasmic Pon and uniformly cortical and strongly spindle-associated Miranda/Prospero; the apical proteins Insc/Pins are normal or slightly expanded (Fig. 1; Table 1; Pins not shown). Although *lgl^{GLC}* and *dlg^{GLC}* embryos show striking defects in neuroblast basal protein localization, they also show an early loss of embryonic epithelial apical/basal polarity¹² (data not shown), which could indirectly cause the observed neuroblast defects.

To determine the neuroblast-specific function of Lgl and Dlg, we scored Lgl- or Dlg-depleted neuroblasts in embryos or larvae where epithelial development occurs normally. First, we assay homozygous null *lgl⁴* embryos, in which maternal Lgl protein allows normal embryonic epithelial development^{12,13} (including Armadillo,

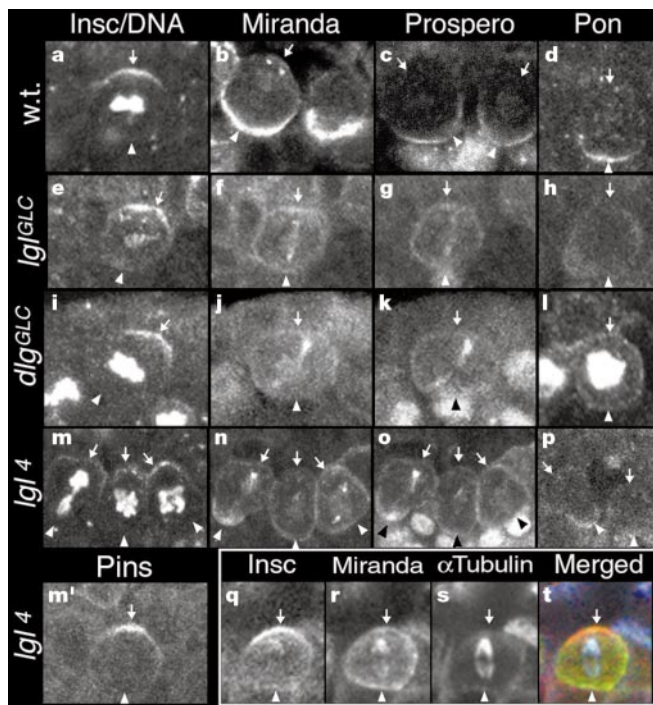


Figure 1 Lgl and Dlg are required for basal protein targeting but not spindle orientation in embryonic neuroblasts. The figure shows wild-type stage 10 embryos (a–d), *lgl^{GLC}* stage 10 embryos (e–h), *dlg^{GLC}* stage 10 embryos (i–l) and *lgl⁴* stage 16–17 embryos (m–p); proteins assayed are indicated at top; mitotic neuroblasts are labelled (apical side, arrow; basal side, arrowhead); cell-cycle staging was done with the mitotic DNA marker phosphohistone H3 (PH3; a, e, i, l, m; not shown in other panels). The apical markers Insc (a, e, i, m; co-labelled with DNA) and Pins (m' and data not shown) are normal in all genotypes; the basal markers Miranda (b, f, j, n), Prospero (c, g, k, o) and Pon (d, h, l, p; co-labelled with DNA in l) are delocalized in *lgl^{GLC}*, *dlg^{GLC}* and *lgl⁴* embryos. Miranda/Prospero are spindle/centrosome-associated whereas Pon is cytoplasmic. 'Rescue' of basal crescents is often seen at late anaphase and telophase (left neuroblast in n, o, p). Spindle orientation is normal in *lgl⁴* embryos. q, Insc is apical; r, Miranda is delocalized onto spindle; s, the mitotic spindle is apical/basal oriented (α -tubulin staining); t, merged image.

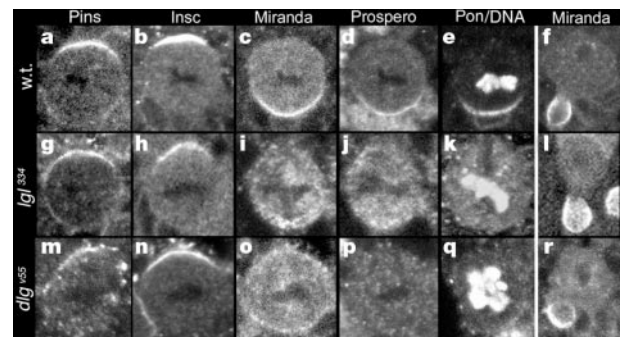


Figure 2 Lgl and Dlg are required for basal protein targeting in larval brain neuroblasts. The figure shows wild-type metaphase (a–e) and telophase (f) neuroblasts; *lgl³³⁴* metaphase (g–k) and telophase (l) neuroblasts; and *dlg³³⁴* metaphase (m–q) and telophase (r) neuroblasts stained for the indicated apical proteins (Insc and Pins), basal proteins (Miranda, Prospero and Pon), and mitotic DNA marker phosphohistone H3 (shown only in Pon column). *lgl³³⁴* and *dlg³³⁴* neuroblasts form normal Insc/Pins crescents, but Miranda/Prospero/Pon are cytoplasmic, uniform cortical, or weakly centrosome/spindle-associated (Miranda/Prospero); telophase neuroblasts often show normal basal protein localization (f, l, r). Dark area in the neuroblast cytoplasm is site of condensed DNA.

Table 1 Lgl and Dlg are required for basal protein targeting

	Apical Insc (metaphase)	Basal Miranda (metaphase)	Basal Miranda (telophase)
Wild type	89% (121)	93% (121)	100% (75)
<i>lgl⁴</i> (stage 10)	85% (204)	37% (94)	83% (97)
<i>lgl⁴</i> (stage 17)	87% (97)	0% (97)	64% (84)
<i>lgl^{GLC}</i> (stage 10)	65% (86)	0% (144)	57% (51)
<i>dlg^{GLC}</i> (stage 10)	74% (126)	0% (126)	68% (22)
<i>dlg⁵⁵</i> (larvae)	100% (7)	0% (27)	100% (5)

Crumbs and Dlg localization; data not shown). In stage 16–17 *lgl⁴* embryos, mitotic neuroblasts show normal Baz/Insc/Pins apical crescents (Fig. 1m, m'; Baz, not shown; Table 1), and normal spindle orientation (Fig. 1q–t), but Miranda/Prospero are delocalized onto the spindle and around the cortex and Pon is cytoplasmic (Fig. 1n–p). This phenotype is less severe in early embryos but fully penetrant in older embryos (Table 1), presumably due to progressive loss of maternal Lgl protein. Second, we assayed neuroblasts in *lgl³³⁴* or

dlg^{v55} homozygous larvae—these larval neuroblasts are persistent embryonic neuroblasts¹⁵ that develop from a normal embryonic epithelium due to maternal Lgl and Dlg protein function. Wild-type larval metaphase neuroblasts have Insc/Pins crescents opposite that of Miranda/Prospero/Pon/Numb crescents (Fig. 2a–e), whereas homozygous *lgl³³⁴* or *dlg^{v55}* larval metaphase neuroblasts show normal Insc/Pins crescents but Miranda/Prospero/Pon proteins are cytoplasmic, uniformly cortical (Fig. 2g–k, m–q), and weakly spindle-associated (data not shown). We conclude that Lgl and Dlg are required specifically in neuroblasts for basal protein targeting, without affecting apical protein localization or spindle orientation.

The *lgl* and *dlg* neuroblast basal localization phenotype is cell-cycle dependent. *lgl* and *dlg* mutant embryonic and larval neuroblasts show a fully penetrant loss of basal protein targeting at metaphase, but by late anaphase or telophase most neuroblasts show normal basal protein localization (Fig. 1n–p, Fig. 2f, l, r; Table 1); ‘telophase rescue’ of basal protein localization also occurs in *baz*, *insc* and *pon* mutants^{5,6,16}. These results indicate that there are probably multiple mechanisms for basal protein localization.

Delocalization of the Prospero and Numb proteins produces defects in the nervous system and other tissues^{16–18}, so we scored *lgl* mutant embryos for cell fate defects. *lgl^{GLC}* embryos have severe morphological defects which preclude analysis, and *lgl⁴* embryos can only be scored for late embryonic phenotypes, due to persistence of maternal Lgl protein. We find that *lgl⁴* embryos show a decrease in Even-skipped lateral (EL) neuron number at stage 17 (wild type: 100% abdominal hemisegments $\geq$ 8 ELs, $n = 133$; *lgl⁴* embryos: 27% abdominal hemisegments $\geq$ 8 ELs, $n = 29$). A similar

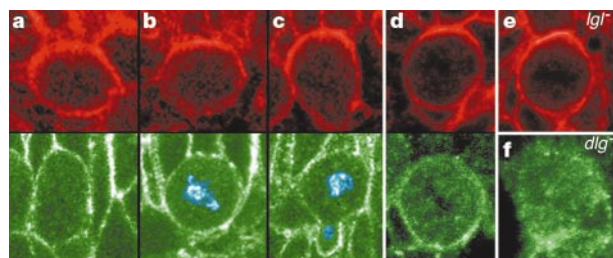


Figure 3 Dlg and Lgl protein localization in neuroblasts. Dlg (red), Lgl (green), phosphohistone H3 mitotic DNA marker (blue pseudocolour in **a–c**, lower panels); apical up. Dlg is uniformly cortical plus apically enriched in wild-type embryonic neuroblasts at interphase (**a**), metaphase (**b**) and telophase (**c**); in larval neuroblasts (**d**); as well as in *lgl³³⁴* larval neuroblasts (**e**). Lgl is uniformly cortical plus cytoplasmic in wild-type embryonic neuroblasts at interphase (**a**), metaphase (**b**) and telophase (**c**); in larval neuroblasts (**d**); but is dissociated from the cortex in *dlg^{v55}* larval neuroblasts (**f**).

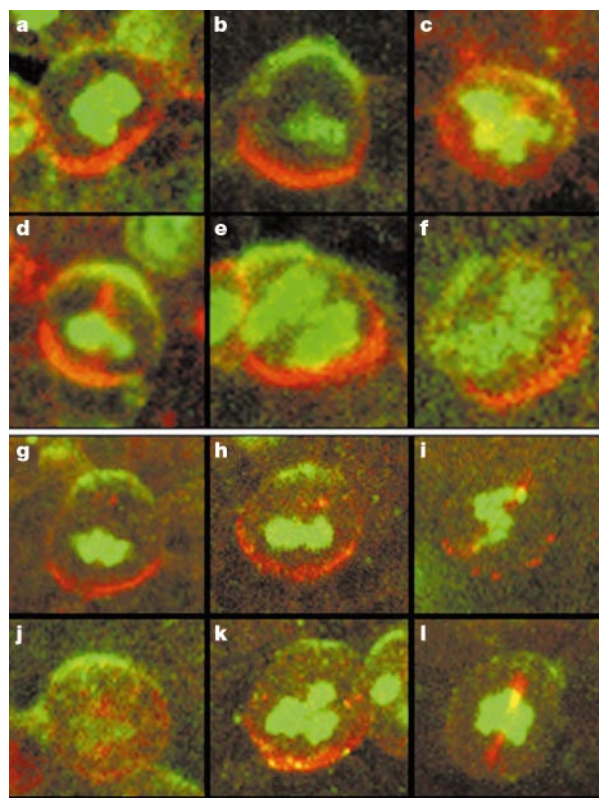
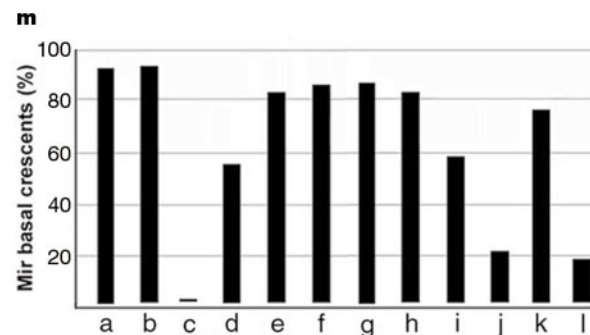


Figure 4 Myosins negatively and positively regulate basal protein targeting in neuroblasts. A single neuroblast is shown triple-labelled for Miranda (red), Insc (green), and DNA (phosphohistone H3; green); apical up. **a–f**, Reduction of zygotic myosin II (encoded by *zip*) suppresses the zygotic *lgl* phenotype in stage 17 embryos. **a**, Wild type: Insc apical, Miranda basal; **b**, *zip¹ / zip¹*: Insc apical, Miranda basal; **c**, *lgl⁴ / lgl⁴*: Insc apical, Miranda delocalized; **d**, *lgl⁴ zip¹ / lgl⁴ +*: Insc apical, Miranda basal and spindle associated; **e**, *lgl⁴ zip¹ / zip¹*: Insc apical, Miranda basal, neuroblast polyploid due to loss of *zip*; **f**, *lgl⁴ zip¹ / lgl⁴ zip¹*: Insc apical, Miranda basal, neuroblast polyploid due to loss of



zip. g–l, The general myosin inhibitor BDM affects basal protein targeting in stage 10 embryos. **g**, Wild type, 0 mM BDM: Insc apical, Miranda basal. **h**, Wild type, 25 mM BDM: Insc apical, Miranda basal. **i**, Wild type, 50 mM BDM: Insc weakly apical, Miranda delocalized and aggregated. **j**, *lgl⁴ / lgl⁴*, 0 mM BDM: Insc apical, Miranda delocalized onto spindle. **k**, *lgl⁴ / lgl⁴*, 25 mM BDM: Insc apical, Miranda basal and aggregated. **l**, *lgl⁴ / lgl⁴*, 50 mM BDM: Insc weakly apical, Miranda delocalized, aggregated, and spindle-associated. **m**, Quantitation of data; letters refer to treatments described in the indicated panels (**a–l**). For all experiments, $n > 150$ neuroblasts scored.

but stronger phenotype is seen in *numb* mutants¹⁹, suggesting that the *lgl* phenotype may be due to delocalization of Numb during the GMC divisions that produce the EL neurons. The relatively mild *lgl* phenotype could be due to 'telophase rescue' of Numb protein in these GMCs, or to maternal Lgl protein.

In wild-type embryonic neuroblasts, Dlg protein is cortical with an apical crescent from late interphase to the end of mitosis (Fig. 3a–c, top panels) that co-localizes with Baz/Insc/Pins (data not shown), whereas Lgl protein is uniformly cortical and weakly cytoplasmic (Fig. 3a–c, bottom panels). Similarly, larval neuroblasts show apical Dlg and uniform cortical/cytoplasmic Lgl localization (Fig. 3d). *lgl* mutants show normal Dlg localization (Fig. 3e), but *dlg* mutants show loss of cortical Lgl protein (Fig. 3f). Thus, Dlg acts upstream of Lgl for its localization, but not necessarily its function.

How does Lgl regulate basal protein targeting? Lgl binds non-muscle myosin II in all organisms tested^{8,11,20}, and *sro7/77* and *myo1* (encoding Lgl-related proteins and myosin II, respectively) show strong negative genetic interactions in yeast¹¹. We tested for genetic interactions between *lgl*¹ and two different null mutations in *zipper* (encoding myosin II), scoring Miranda basal localization in stage 17 neuroblasts, when maternal Lgl and Myosin II protein levels are lowest. Wild-type and *zip*¹ embryos have normal basal protein localization, whereas *lgl*¹ embryos show complete delocalization of basal proteins (Fig. 4a–c). However, *lgl*¹ embryos lacking one copy of myosin II (*lgl*¹ *zip*¹ / *lgl*¹ +) show a significant increase in basal protein targeting (Fig. 4d); and *lgl*¹ *zip*¹ or *lgl*¹ *zip*^{1X62} double mutant embryos show virtually normal basal protein targeting (Fig. 4e, f). Thus, reducing myosin II levels strongly suppresses the *lgl* phenotype, indicating that myosin II can inhibit basal targeting when Lgl levels are low.

In addition, the general myosin inhibitor 2,3-butanedione monoxime (BDM)²¹ can suppress the *lgl* phenotype: stage 10 *lgl*¹ embryos treated with 25 mM BDM show a significant increase in basal protein localization (Fig. 4k) compared with sham-treated stage 10 *lgl*¹ embryos (Fig. 4j); 25 mM BDM has no effect on wild-type embryos (compare Fig. 4g and h). In contrast, wild-type or *lgl*¹ embryos treated with 50 mM BDM show delocalization of Miranda, Prospero (Fig. 4i, l) and Pon²². These data, quantified in Fig. 4m, indicate that a myosin which is sensitive to 25 mM BDM inhibits basal protein localization in *lgl* embryos (probably myosin II), and at least one myosin that is sensitive to 50 mM BDM promotes basal protein targeting in mitotic neuroblasts.

Here we show that in neuroblasts, Lgl and Dlg regulate targeting of all known basal proteins without affecting apical protein localization or spindle orientation. In epithelia, Lgl and Dlg are necessary to restrict proteins to the apical membrane domain¹². Lgl could promote protein targeting to specific membrane domains in both neuroblasts (basal) and epithelia (apical), similar to the role of Lgl-related proteins in facilitating secretory vesicle fusion at specific membrane domains in yeast and mammals^{9,10}. If so, Lgl must act in neuroblasts via a secretory pathway that is independent of brefeldin A, because we show that treatment with brefeldin A disrupts Golgi, inhibits Wingless secretion, but does not block basal protein targeting (see Supplementary Information Fig. 1i–n). Alternatively, Lgl may actively promote actomyosin-dependent localization of basal proteins and/or function to keep myosin II levels low so that they do not interfere with myosin-dependent basal localization. A general function of the Lgl protein family may be to increase the fidelity of protein targeting to specific domains of the plasma membrane. □

Methods

Fly strains and genetics

All fly strains were raised at 25 °C on standard media. *lgl*¹, *lgl*³³⁴, *zip*¹ and *zip*^{1X62} are all strong or null alleles; *dlg*⁵⁵ is a hypomorph. *lgl*¹ or *dlg*⁵⁵ germline clones were generated using FLP-DFS methods²³ and crossed to *lgl*¹ / CyO *ftz-lacZ* or Y / FM7 *ftz-lacZ* males to generate β-galactosidase-negative *lgl*¹^{GLC} or *dlg*⁵⁵^{GLC} embryos. Females with *lgl*¹ germline

clones were crossed to *lgl*¹ *zip*¹ / CyO *ftz-lacZ* males to generate β-galactosidase-negative embryos having no Lgl function and 50% myosin II function. The *lgl*¹ *zip*¹ and *lgl*¹ *zip*^{1X62} double mutants were generated by recombination.

Drug treatments

The myosin inhibitor 2,3-butanedione monoxime (Sigma) was dissolved directly in Schneider's medium (SM) at 25 or 50 mM and used within 1 h. Brefeldin A (Sigma) was made as 10 mg ml⁻¹ methanol stock and diluted in SM at 40 μg ml⁻¹. Drug treatments were performed for 30 min as previously described²².

Antibody generation, staining, imaging and analysis

Antibody stainings were performed as previously described^{3,24}. Primary antibodies were rat anti-Miranda (1:200; generated against the peptide SPPQKQVLKARNI), rat anti-Lgl (1:500; generated against the peptide HEKTNKDNGKIGTPKTAPEESQF), rabbit anti-Lgl-C (1:200), mouse anti-Prospero (1:4)³, mouse anti-Eve (1:10), rabbit anti-Inscuteable and rabbit anti-Pins (both 1:1,000), rat anti-Baz (1:200), rabbit anti-Pon (1:1,000), rabbit anti-Numb (1:1,000), mouse anti-Crums (1:20), guinea-pig anti-Dlg (1:2,000), rabbit anti-β-galactosidase (1:5,000; Cappel), rabbit anti-phosphohistone H3 (PH3; 1:5,000; Upstate Biotech.), mouse anti-Armadillo (1:100), rabbit anti-Wingless (1:50), and mouse anti-Golgi²⁵ (1:200). Imaging was done on a Biorad 1024 confocal microscope.

Received 11 August; accepted 6 November 2000.

1. Doe, C. Q. & Bowerman, B. Asymmetric cell division: fly neuroblast meets worm zygote. *Curr. Opin. Cell Biol.* (in the press).
2. Jan, Y. N. & Jan, L. Y. Asymmetry across species. *Nature Cell Biol.* **1**, E42–E44 (1999).
3. Spana, E. P. & Doe, C. Q. The prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
4. Schaefer, M., Shevchenko, A. & Knoblich, J. A. A protein complex containing Inscuteable and the Go-binding protein Pins orients asymmetric cell divisions in *Drosophila*. *Curr. Biol.* **10**, 353–362 (2000).
5. Schober, M., Schaefer, M. & Knoblich, J. A. Bazooka recruits Inscuteable to orient asymmetric cell divisions in *Drosophila* neuroblasts. *Nature* **402**, 548–551 (1999).
6. Wodarz, A., Ramrath, A., Kuchinke, U. & Knust, E. Bazooka provides an apical cue for Inscuteable localization in *Drosophila* neuroblasts. *Nature* **402**, 544–547 (1999).
7. Yu, E., Morin, X., Cai, Y., Yang, X. & Chia, W. Analysis of partner of inscuteable, a novel player of *Drosophila* asymmetric divisions, reveals two distinct steps in inscuteable apical localization. *Cell* **100**, 399–409 (2000).
8. Strand, D. et al. A human homologue of the *Drosophila* tumour suppressor gene *l(2)gl* maps to 17p11.2–12 and codes for a cytoskeletal protein that associates with nonmuscle myosin II heavy chain. *Oncogene* **11**, 291–301 (1995).
9. Fujita, Y. et al. Tomosyn: a syntaxin-1-binding protein that forms a novel complex in the neurotransmitter release process. *Neuron* **20**, 905–915 (1998).
10. Lehman, K., Rossi, G., Adamo, J. E. & Brennwald, P. Yeast homologues of tomosyn and lethal giant larvae function in exocytosis and are associated with the plasma membrane SNARE, Sec9. *J. Cell. Biol.* **146**, 125–140 (1999).
11. Kagami, M., Toh-e, A. & Matsui, Y. Sro7p, a *Saccharomyces cerevisiae* counterpart of the tumor suppressor *l(2)gl* protein, is related to myosins in function. *Genetics* **149**, 1717–1727 (1998).
12. Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. *Science* **289**, 113–116 (2000).
13. Manfruell, P., Arquier, N., Hanratty, W. P. & Semeriva, M. The tumor suppressor gene, *lethal(2)giant larvae* (*l(2)gl*), is required for cell shape change of epithelial cells during *Drosophila* development. *Development* **122**, 2283–2294 (1996).
14. Woods, D. E., Hough, C., Peel, D., Callaini, G. & Bryant, P. J. Dlg protein is required for junction structure, cell polarity, and proliferation control in *Drosophila* epithelia. *J. Cell. Biol.* **134**, 1469–1482 (1996).
15. Truman, J. W. & Bate, M. Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila melanogaster*. *Dev. Biol.* **125**, 145–157 (1988).
16. Lu, B., Rothenberg, M., Jan, L. Y. & Jan, Y. N. Partner of Numb colocalizes with Numb during mitosis and directs Numb asymmetric localization in *Drosophila* neural and muscle progenitors. *Cell* **95**, 225–235 (1998).
17. Ikeshima-Kataoka, H., Skeath, J. B., Nabeshima, Y., Doe, C. Q. & Matsuzaki, F. Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* **390**, 625–629 (1997).
18. Shen, C. P., Jan, L. Y. & Jan, Y. N. Miranda is required for the asymmetric localization of Prospero during mitosis in *Drosophila*. *Cell* **90**, 449–458 (1997).
19. Skeath, J. B. & Doe, C. Q. Sanpodo and Notch act in opposition to Numb to distinguish sibling neuron fates in the *Drosophila* CNS. *Development* **125**, 1857–1865 (1998).
20. Strand, D. et al. The *Drosophila* lethal(2)giant larvae tumor suppressor protein forms homo-oligomers and is associated with nonmuscle myosin II heavy chain. *J. Cell. Biol.* **127**, 1361–1373 (1994).
21. Hajjar, R. J., Ingwall, J. S. & Gwathmey, J. K. Mechanism of action of 2,3-butanedione monoxime on contracture during metabolic inhibition. *Am. J. Physiol.* **267**, H100–H108 (1994).
22. Lu, B., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of Numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
23. Chou, T. B. & Perrimon, N. The autosomal FLP-DFS technique for generating germline mosaics in *Drosophila melanogaster*. *Genetics* **144**, 1673–1679 (1996).
24. Manning, L. & Doe, C. Q. Prospero distinguishes sibling cell fate without asymmetric localization in the *Drosophila* adult external sense organ lineage. *Development* **126**, 2063–2071 (1999).
25. Stanley, H., Botas, J. & Malhotra, V. The mechanism of Golgi segregation during mitosis is cell type-specific. *Proc. Natl Acad. Sci. USA* **94**, 14467–14470 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank H. Dou for assistance in the deficiency screen; S. Scheider for phylogenetic analysis of Lgl; S. Siegrist and S. Fuerstenberg for Supplementary Fig. 1a–h; D. Bilder, J. K. Roy and N. Perrimon for sharing unpublished results, the *FRT¹lgl¹* chromosome, and some *lgl^{GLC}* embryos; B. Mechler, V. Malhotra, B. Chia, E. Knust, M. Peifer, Y. N. Jan, P. Bryant, C. Logan, R. Nusse and the Bloomington stock centre for providing antibodies or fly stocks; and T. Isshiki, S. Fuerstenberg, B. Bowerman for comments and sharing unpublished data. This work was supported by the NIH and HHMI, of which C.Q.D. is an Associate Investigator.

Correspondence and requests for materials should be addressed to C.Q.D. (e-mail: cdoe@uoneuro.uoregon.edu).

T-cell-mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN- γ

Hiroshi Takayanagi¹*, Kouetsu Ogasawara¹*, Shigeaki Hida¹*, Tomoki Chiba¹*, Shigeo Murata¹*, Kojiro Sato¹*, Akinori Takaoka¹*, Taeko Yokochi¹*, Hiromi Oda¹*, Keiji Tanaka¹*, Kozo Nakamura¹† & Tadatsugu Taniguchi¹*

^{*} Department of Immunology, and [†] Orthopaedic Surgery, Faculty of Medicine and Graduate School of Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[‡] Department of Molecular Oncology, Tokyo Metropolitan Institute of Medical Science, and CREST, Japan Technology Corporation, 3-18-22, Honkomagome, Bunkyo-ku, Tokyo 113-8613, Japan

Bone resorption is regulated by the immune system^{1,2}, where T-cell expression of RANKL (receptor activator of nuclear factor (NF)- κ B ligand), a member of the tumour-necrosis factor family that is essential for osteoclastogenesis, may contribute to pathological conditions, such as autoimmune arthritis^{3,4}. However, whether activated T cells maintain bone homeostasis by counterbalancing the action of RANKL remains unknown. Here we show that T-cell production of interferon (IFN)- γ strongly suppresses osteoclastogenesis by interfering with the RANKL–RANK signalling pathway. IFN- γ induces rapid degradation of the RANK adapter protein, TRAF6 (tumour necrosis factor receptor-associated factor 6), which results in strong inhibition of the RANKL-induced activation of the transcription factor NF- κ B and JNK. This inhibition of osteoclastogenesis is rescued by over-expressing TRAF6 in precursor cells, which indicates that TRAF6 is the target critical for the IFN- γ action. Furthermore, we provide evidence that the accelerated degradation of TRAF6 requires both its ubiquitination, which is initiated by RANKL, and IFN- γ -induced activation of the ubiquitin–proteasome system. Our study shows that there is cross-talk between the tumour necrosis factor and IFN families of cytokines, through which IFN- γ provides a negative link between T-cell activation and bone resorption. Our results may offer a therapeutic approach to treat the inflammation-induced tissue breakdown.

Immune response is essential for host defence, but its prolonged or aberrant activation under particular autoimmune conditions often results in tissue destruction mediated by effector cells. In autoimmune arthritis, enhanced-osteoclastic bone resorption causes severe bone damage leading to progressive joint destruction^{4,5}, in which T-cell expression of RANKL (also called TRANCE/ODF/OPGL) may have a critical function. In fact, activated T cells promote osteoclastogenesis through RANKL expression *in vitro*^{4,6}. In addition to RANKL, other immunomodulatory factors that are expressed by T cells also modulate osteoclastogenesis^{1,2,7,8};

however, the molecular mechanism of the regulation remains unknown.

Using an established model of endotoxin-induced bone resorption, in which the essential involvement of T cells has been documented^{9,10}, we observed a notable exacerbation of osteoclast formation and bone destruction in mice lacking one of the IFN- γ receptor (IFN- γ R) components, IFNGR1 (ref. 11) (IFN- γ R^{−/−} mice; Fig. 1a). This observation indicates that IFN- γ may be a critical factor in maintaining bone integrity through a T-cell-mediated process. It is also consistent with enhanced severity in the collagen-induced model of T-cell-mediated autoimmune arthritis, which is

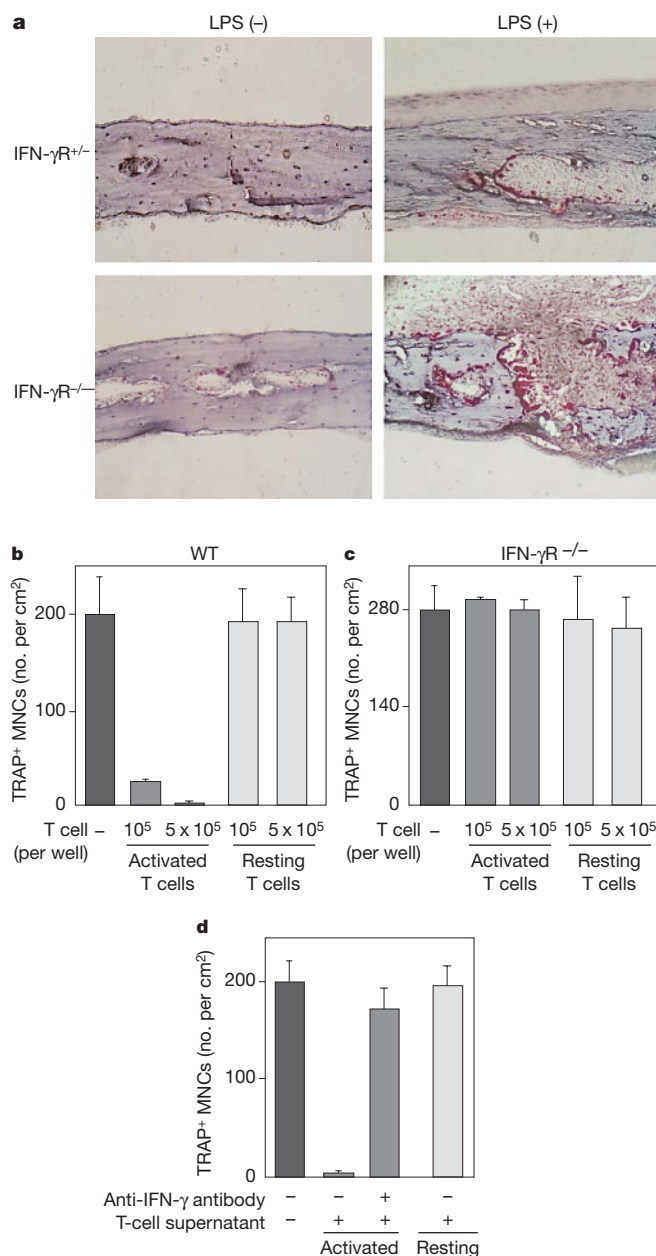


Figure 1 T-cell-mediated regulation of osteoclastogenesis by IFN- γ . **a**, Histology of calvarial bone injected with saline or lipopolysaccharide in IFN- γ R^{+/−} and IFN- γ R^{−/−} mice (TRAP and haematoxylin staining). The bone morphometric analysis indicates more than a twofold increase in the number of osteoclast (TRAP⁺ MNCs), and in the eroded surface of the calvarial bone in IFN- γ R^{−/−} mice. **b**, Suppression of osteoclastogenesis by activated T cells. Activated splenic T cells or resting T cells were co-cultured with RANKL/M-CSF-stimulated BMMs. **c**, Cancellation of T-cell-mediated suppression of osteoclastogenesis in co-culture with IFN- γ R^{−/−} BMMs. **d**, Effect of T-cell supernatant on RANKL-induced osteoclastogenesis in the absence or presence of anti-IFN- γ antibody.

seen in IFN- γ ^R mice^{12,13}.

We examined the effect of IFN- γ -producing T cells on osteoclastogenesis, by *in vitro* co-culture of bone-marrow-derived monocyte/macrophage precursor cells (BMMs) stimulated by RANKL in

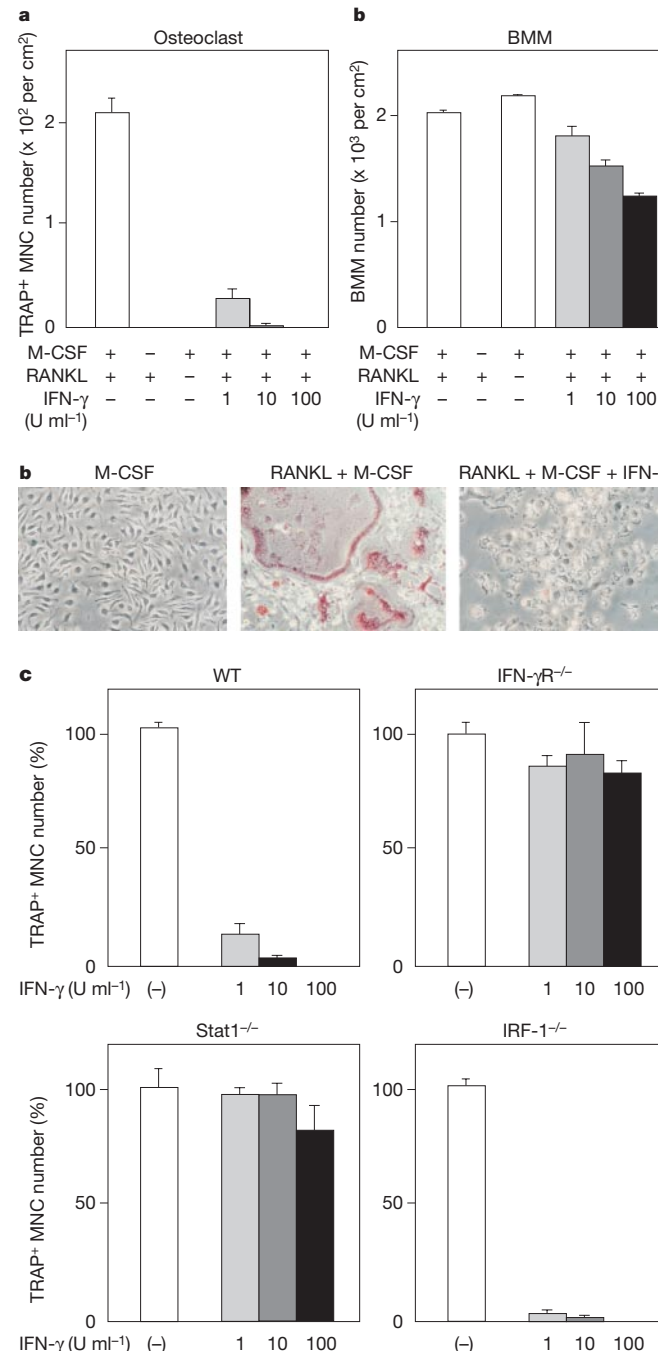


Figure 2 Inhibition of RANKL-induced osteoclastogenesis by IFN- γ through Stat1 activation pathway. **a**, Effect of IFN- γ on RANKL-induced osteoclastogenesis and BMM proliferation. IFN- γ has a marked inhibitory effect on osteoclast formation in RANKL/M-CSF-stimulated BMMs, and a marginal effect on the number of BMMs.

b, Microphotograph of *in vitro* osteoclast formation system (TRAP staining). Left, BMMs cultured with M-CSF alone. Middle, RANKL/M-CSF-stimulated BMMs efficiently differentiated into TRAP⁺ MNCs. Right, Inhibition of RANKL/M-CSF-stimulated osteoclastogenesis by IFN- γ (100 U ml⁻¹). **c**, Requirement of IFN- γ R and Stat1 for IFN- γ inhibition of osteoclastogenesis. Effect of IFN- γ on osteoclastogenesis in RANKL/M-CSF-stimulated BMMs derived from mice lacking IFN- γ R, Stat1 or IRF-1. The values for 100% are wild type, 200; IFN- γ ^R, 280; Stat1^{-/-}, 280; and IRF-1^{-/-}, 240 (TRAP⁺ MNCs per cm²).

the presence of macrophage colony-stimulating factor (M-CSF), with activated or resting splenic T cells. Osteoclast formation was strongly inhibited by the co-culture with T cells activated with the anti-CD3 antibody, but not with resting T cells (Fig. 1b). Under these conditions, most of the BMMs underwent differentiation into mature macrophages, and similar results were obtained by co-culture with RANKL-stimulated cells of the macrophage cell line RAW 264.7 (data not shown). When activated T cells were co-cultured with IFN- γ ^R BMMs stimulated by RANKL, the inhibitory effect of activated T cells was completely cancelled (Fig. 1c). These IFN- γ ^R-BMMs differentiated into osteoclasts, albeit with low efficiency, when co-cultured with activated T cells even in the absence of recombinant RANKL (data not shown), suggesting that the effect of T cells on osteoclastogenesis depends on the balance between RANKL and IFN- γ . As expected, activated T-cell supernatant suppressed osteoclastogenesis induced by recombinant RANKL, and this suppressive activity was inhibited by a neutralizing antibody against IFN- γ (Fig. 1d). These results indicate that IFN- γ may be critical for T-cell suppression of RANKL-induced osteoclastogenesis.

A marked inhibitory effect of IFN- γ was observed by an *in vitro* assay for osteoclast formation in RANKL-stimulated BMMs (Fig. 2a, b). A low level of IFN- γ strongly inhibits osteoclast formation, even in the presence of an excess amount of RANKL, whereas IFN- γ had a relatively marginal effect on the survival and proliferation of BMMs, which are dependent on M-CSF but not on RANKL (Fig. 2a, b). Furthermore, IFN- γ inhibited RANKL-induced osteoclastogenesis of RAW 264.7 cells, which is M-CSF-independent (data not shown). These results suggest that IFN- γ interferes selectively with the RANKL signalling pathway. IFN- γ signals the cell through activation of the transcription factor, Stat1 (ref. 14). The active form of Stat1, called IFN- γ -activated factor (GAF), induces target genes of IFN- γ either directly or through the induction of the transcription factor, IRF-1 (ref. 15). The inhibitory effect of IFN- γ on osteoclastogenesis was cancelled in IFN- γ ^R and Stat1^{-/-} mice, but not in IRF-1^{-/-} mice (Fig. 2c). Thus, the Stat1 (GAF)-mediated IRF-1-independent gene induction pathway is critical for interfering with the RANKL signalling pathway.

RANKL stimulates its receptor, RANK, which recruits adapter proteins of the TRAF family, thereby leading to activation of the NF- κ B and Jun N-terminal kinase (JNK) pathways¹⁶. Among TRAF family members TRAF6 is the most critical in RANKL signalling, which is evident with the osteopetrotic phenotype owing to the absence of bone-resorbing osteoclasts in TRAF6^{-/-} mice^{17,18}. To investigate the target of IFN- γ inhibition of osteoclastogenesis, we examined RANKL-induced activation of NF- κ B and JNK in IFN- γ -treated BMMs. An electrophoretic mobility shift assay (EMSA) showed that RANKL-induced activation of NF- κ B was markedly inhibited in IFN- γ -treated BMMs (Fig. 3a). Furthermore, RANKL-induced activation of JNK was also inhibited by IFN- γ (Fig. 3b). Immunoblot analysis of downstream effector molecules of the RANKL signalling pathway showed that the expression of TRAF6 is selectively and markedly inhibited by IFN- γ in BMMs stimulated by RANKL and M-CSF (Fig. 3c).

To assess further whether inhibition of TRAF6 expression contributes to the IFN- γ -mediated suppression of osteoclastogenesis, we constructed a retrovirus vector carrying TRAF6 complementary DNA with an internal ribosomal entry site-enhanced green fluorescent protein (pMX-TRAF6-IRES-EGFP), and then infected the BMMs with the virus. Flowcytometric and immunoblot analyses revealed an efficient gene transduction into BMMs, by the TRAF6-expressing virus and a control virus expressing GFP only (pMX-IRES-EGFP) (Fig. 3d). Notably, overexpression of TRAF6 in BMMs rendered them resistant to IFN- γ -mediated inhibition of osteoclastogenesis (Fig. 3e). In these cells, RANKL-induced activation of NF- κ B and JNK was also restored (data not shown), indicating further that TRAF6, rather than these downstream effector molecules, is the

critical target of the IFN- γ action. We note that about 50% of the TRAF6-GFP-positive cells differentiated into tartrate-resistant acid-phosphatase-positive multinucleated cells (TRAP⁺ MNCs), even in the presence of IFN- γ (Fig. 3e). These differentiated cells expressed calcitonin receptors and exhibited a bone-resorbing activity on dentine slices (data not shown). These results show that downregulation of TRAF6 expression accounts, at least in part, for the IFN- γ inhibition of osteoclastogenesis. Our results are more consistent with one of the two reports on TRAF6^{-/-} mice^{17,18}, suggesting that TRAF6 is critical for osteoclast differentiation.

We investigated further the mechanism by which TRAF6

expression is downregulated by IFN- γ . RANKL stimulation of RAW 264.7 cells caused a slight augmentation of TRAF6 messenger RNA expression, and this expression profile remained unaffected after the addition of IFN- γ (Fig. 4a). This observation led us to examine TRAF6 protein levels after RANKL stimulation in the presence or absence of IFN- γ by pulse-chase experiments. Notably, stimulation by RANKL *per se* promoted TRAF6 degradation, which was markedly accelerated by IFN- γ (Fig. 4b). As the TRAF6 level is upregulated during stimulation by RANKL (Fig. 3c), the equilibrium of TRAF6 protein seems to be shifted toward its synthesis rather than its degradation in the absence of IFN- γ ; but the addition

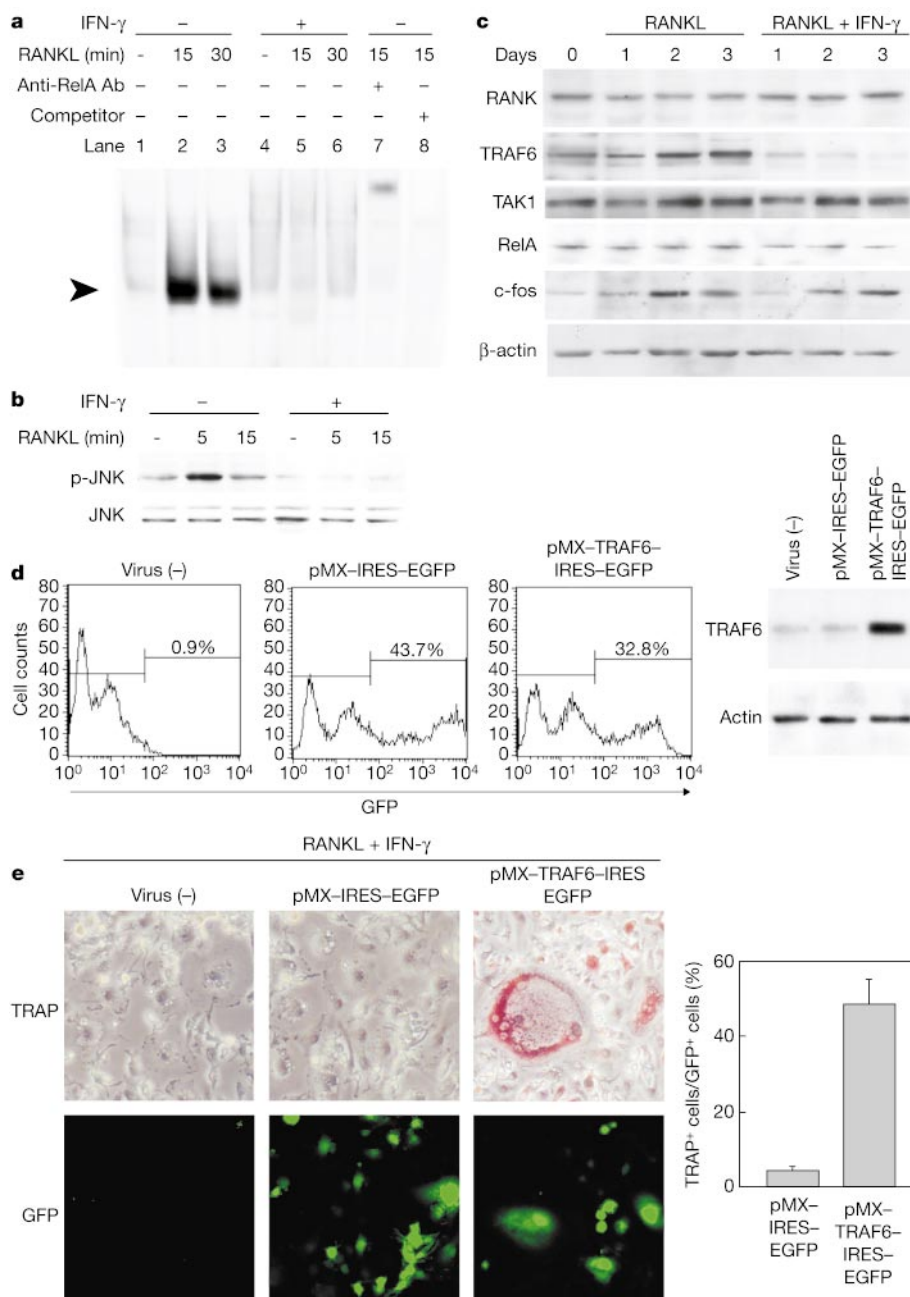


Figure 3 Inhibition of RANKL signalling by IFN- γ through downregulation of TRAF6. **a**, EMSA for NF- κ B activity. The NF- κ B complex (arrowhead) formation peaked 15 min after stimulation (lanes 1–3). It is supershifted by an anti-RelA antibody (lane 7), and inhibited by the unlabelled probe (lane 8). This complex was barely formed in IFN- γ -treated BMMs (lanes 4–6). **b**, Inhibition of JNK activation by IFN- γ . **c**, Effect of IFN- γ on the expression of downstream effector molecules involved in the RANKL signalling. Fold increase in TRAF6 protein normalized by actin is as follows: RANKL, 2.23, 2.93 and 3.84;

RANKL + IFN- γ , 0.29, 0.18 and 0.07. **d**, Efficiency of retroviral gene transfer into primary BMMs. GFP expression by flowcytometry (left) and TRAF6 immunoblotting (right) are shown. **e**, Effect of TRAF6 overexpression on IFN- γ inhibition of osteoclastogenesis. GFP⁺ cells were identified under a fluorescence microscope, and cells were stained for TRAP (left). More than 50% of TRAF6 virus-infected BMMs could differentiate into TRAP⁺ MNCs (right).

of IFN- γ reverses the equilibrium. IFN- γ alone had no effect on TRAF6 expression (data not shown), however, suggesting that accelerated TRAF6 degradation by IFN- γ requires RANKL signaling. TRAF6 degradation is inhibited by proteasome inhibitors, such as MG132 and lactacystin, and by cycloheximide (Fig. 4c); an observation that prompted us to examine further the involvement of the ubiquitin–proteasome system.

The proteasome is a multi-subunit protease complex that destroys a number of cellular proteins, which acquire a specific degradation signal, such as a multi-ubiquitin chain¹⁹. To determine whether TRAF6 is a direct target for the ubiquitin–proteasome system, haemagglutinin (HA)-tagged ubiquitin cDNA was transiently overexpressed in RAW 264.7 cells. Polyubiquitinated TRAF6 was detected in RANKL-stimulated cells, and this RANKL-dependent

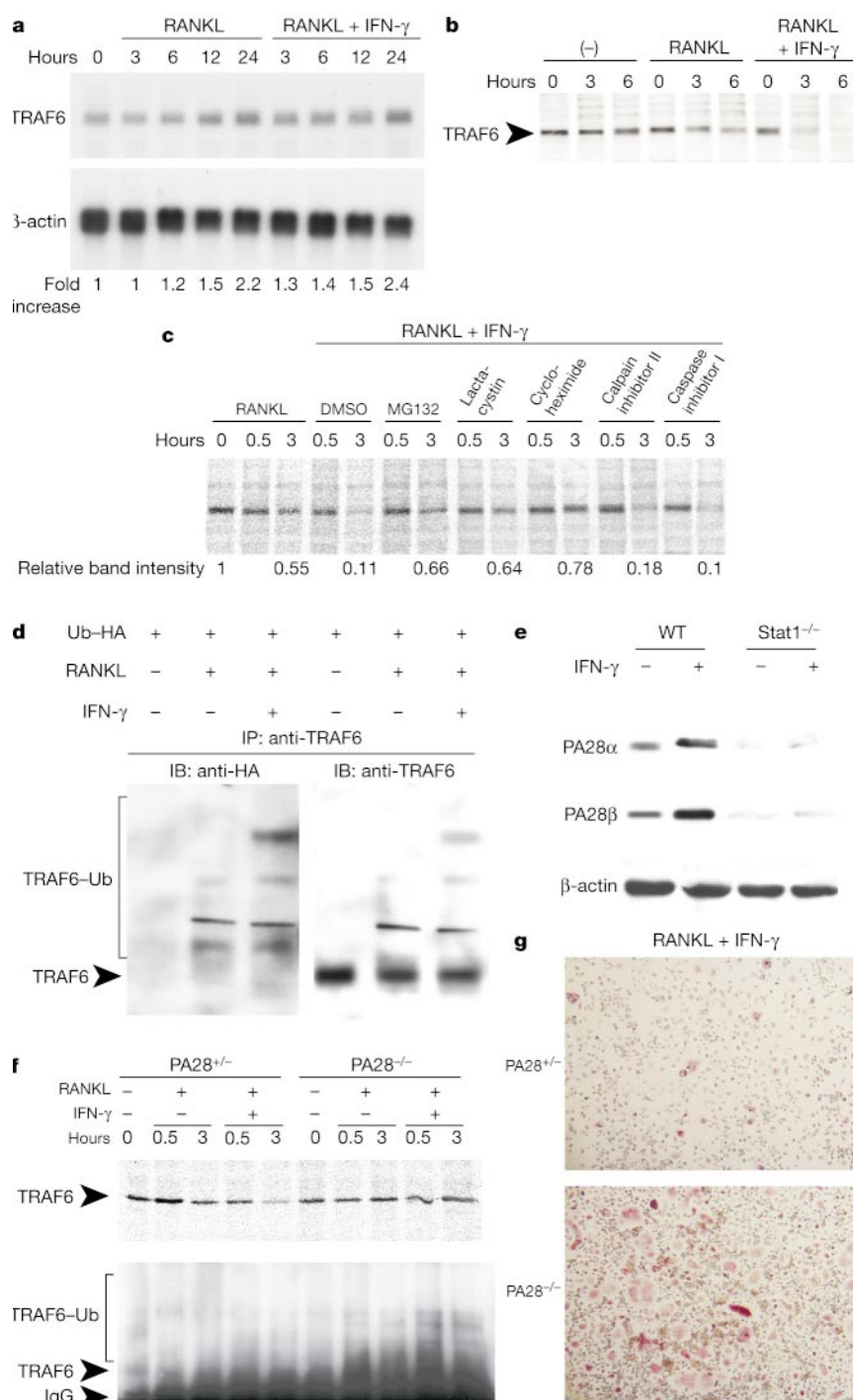


Figure 4 Involvement of the ubiquitin–proteasome pathway in the acceleration of TRAF6 proteolysis. **a**, Effect of IFN- γ on TRAF6 mRNA expression (RNA blotting). **b**, Analysis of the TRAF6 turnover by pulse-chase experiments. **c**, Pulse-chase experiments in the presence of protease inhibitors or cycloheximide. **d**, Ubiquitination of TRAF6. Immunoprecipitates with an anti-TRAF6 polyclonal antibody was blotted with an anti-HA antibody (left), or with an anti-TRAF6 monoclonal antibody (right). **e**, Stat1-dependent

induction of PA28 α and PA28 β by IFN- γ (immunoblotting). **f**, Involvement of PA28 in TRAF6 degradation. Pulse-chase analysis for TRAF6 protein in PA28^{+/-} and PA28^{-/-} BMMs (top). The same membrane was blotted with an anti-ubiquitin antibody (bottom). **g**, Microphotograph of *in vitro* osteoclast formation in PA28^{+/-} and PA28^{-/-} BMMs in the presence of RANKL/M-CSF and IFN- γ (1 U ml⁻¹). TRAP⁺ MNC number (%): 14 $\pm$ 3.5% in PA28^{+/-} BMMs versus 75 $\pm$ 8.0% in PA28^{-/-} BMMs.

polyubiquitination was enhanced further by IFN- γ (Fig. 4d).

IFN- γ activates the ubiquitin-proteasome system by inducing the proteasome activator PA28 as well as immunoproteasome components²⁰. PA28 α and PA28 β , which have a vital function in immunoproteasome assembly²¹, also activate the proteasome complex participating in the ATP-dependent proteolysis²². The induction of PA28 α and PA28 β by IFN- γ was not observed in Stat1^{-/-} BMMs, which were resistant to IFN- γ -induced inhibition of osteoclastogenesis (Fig. 4e). To test the involvement of PA28 in the IFN- γ acceleration of TRAF6 degradation, we examined the BMMs from mice lacking both PA28 α and PA28 β generated by gene targeting (S.M., T.C. and K.T., unpublished data). Figure 4f shows that TRAF6 degradation by IFN- γ is suppressed in PA28^{-/-} BMMs. Notably, polyubiquitinated TRAF6, which is hardly detectable in PA28^{+/-} BMMs, can be detected in PA28^{-/-} BMMs even in the absence of proteasome inhibitors. Furthermore, the inhibitory effect of IFN- γ on RANKL-induced osteoclastogenesis is suppressed in PA28^{-/-} BMMs (Fig. 4g). These results indicate that IFN- γ activation of the ubiquitin-proteasome system involving PA28 is critical in the effective degradation of the ubiquitinated TRAF6.

Collectively, inhibition of osteoclastogenesis by IFN- γ involves a unique signalling cross-talk between RANKL and IFN- γ : RANKL induces ubiquitination of TRAF6, which is further enhanced by IFN- γ , and IFN- γ induces the PA28-dependent proteasome activity. Our preliminary results indicate a similar cross-talk between IFN- γ and CD40, which is also known to use TRAF6 for its signalling (data not shown). Although the mechanism by which RANKL stimulation induces ubiquitination of TRAF6 remains unknown, this process may constitute a negative-feedback regulation of the RANKL signalling.

Recently, it has been reported that activated T cells promote osteoclastogenesis through RANKL expression and regulate bone loss in autoimmune arthritis⁴. We have shown that activated T cells can also negatively affect osteoclastogenesis through IFN- γ production. The balance between RANKL and IFN- γ action may regulate the osteoclast formation. For example, during acute immune reactions an enhanced production of IFN- γ counterbalances the augmentation of RANKL expression and reduces aberrant osteoclast formation, as shown by our model of endotoxin-induced bone resorption. In chronic synovitis of rheumatoid arthritis, however, this balance may be skewed in favour of RANKL expression. In this context, despite a significant T-cell infiltration in the arthritic joints IFN- γ expression is suppressed^{23,24}, whereas RANKL expression is enhanced^{25,26}. Thus, the scarcity of IFN- γ and the enhanced expression of RANKL may contribute to the activation of osteoclastogenesis in arthritis. Other factors must be taken into consideration in the control of osteoclastogenesis in arthritis, such as RANKL induction in interleukin-1- or TNF- α -stimulated synovial fibroblasts (ref. 25; and our own unpublished data), and local cytokine production pattern regulated by differentiation status of T cells into T-helper type 1 or 2.

We have provided insight into a new biological function of IFN- γ , which is to protect against calcified tissue destruction on T-cell activation. The disease-limiting effect of IFN- γ in autoimmune arthritis^{12,13} may be explained, at least in part, by its potent suppressive effect on osteoclast development. Although the clinical application of IFN- γ has been difficult, probably because of its disease-promoting effect at the onset phase of autoimmunity²⁷, our results suggest that TRAF6, the critical target of IFN- γ -mediated suppression of osteoclastogenesis, may be a possible molecular target of pharmacological intervention in inflammatory bone diseases. □

Methods

Mice and endotoxin-induced bone resorption

IFN- γ R^{-/-} and IRF-1^{-/-} mice have been described^{11,28}; C57BL/6 mice, with the same background as these mutant mice, served as wild-type (WT) controls. Stat1^{-/-} mice²⁹ were

provided by R. D. Schreiber. We generated mice deficient in both PA28 α and PA28 β by the conventional gene knockout technology (S.M., T.C. and K.T., unpublished data). We administered a local calvarial injection of lipopolysaccharide (Sigma) at 25 mg per kg to 7–8-week-old mice ($n = 10$). After 5 d, we measured osteoclast number per millimetre of trabecular bone surface, and the percentage of bone surface covered by osteoclasts (eroded surface) as described¹⁰.

In vitro osteoclastogenesis

We cultured nonadherent bone marrow cells (5×10^6 per well in a 24-well plate) in α -MEM with 10 ng ml⁻¹ M-CSF (R&D) for 2 d, and used them as BMMs. We cultured these BMMs in the presence of 100 ng ml⁻¹ soluble RANKL (Peprotech) and 10 ng ml⁻¹ M-CSF to generate osteoclasts; we used RANKL and M-CSF at these concentrations in all experiments. We counted TRAP⁺ multinucleated (>3 nuclei) cells 3 d later. All data are expressed as mean $\pm$ s.e.m. ($n = 6$). We characterized TRAP⁺ MNCs as described³⁰, and they satisfied the criteria of authentic osteoclasts, including the bone-resorbing activity on dentine slices and the expression of calcitonin receptors. Contaminated stromal cells do not contribute to this osteoclastogenesis, as BMMs do not form osteoclasts in the absence of soluble RANKL, even if stimulated by potent RANKL-inducers such as 1,25(OH)₂ vitamin D₃.

T-cell activation and co-culture

We purified CD5⁺ T cells using magnetic beads (Miltentiv Biotec) from splenocytes derived from C57BL/6 mice after depletion of adherent cells. We seeded purified T cells (CD3⁺ > 93%) in RPMI 1640 and stimulated them with 5 μ g ml⁻¹ anti-CD3 ϵ antibody (PharMingen) for 1 d. We added T cells, T-cell supernatant, recombinant mouse IFN- γ (Genzyme) or anti-IFN- γ neutralizing antibody (R4-6A2) simultaneously with RANKL into BMM cultures, and after 3 d we evaluated osteoclast formation. We added T-cell supernatant to four times volume of the culture medium. We changed the medium every 48 h.

Activation of NF- κ B and JNK by RANKL

After stimulation with RANKL, cell extract was prepared from BMMs pre-incubated in the presence or absence of IFN- γ for 24 h, and analysed by EMSA using an oligonucleotide probe containing the NF- κ B-binding site of the IFN- β promoter as described²⁸. We performed supershift with an anti-RelA antibody (Santa Cruz). For JNK assay, BMMs pre-incubated in the presence or absence of IFN- γ for 24 h were stimulated with RANKL, and analysed by immunoblotting using an anti-phosphorylated JNK antibody and an anti-JNK antibody (New England Biomed).

Immunoblot and RNA blotting analysis

To evaluate the expression of downstream molecules of RANKL signalling during osteoclastogenesis, we assayed whole-cell lysate of BMMs every 24 h after RANKL addition by immunoblotting using specific antibodies for RANK, TRAF6, TAK1, RelA and c-Fos (Santa Cruz). For detection of PA28, we stimulated BMMs derived from wild-type and Stat1^{-/-} mice with 100 U ml⁻¹ IFN- γ for 24 h, and we prepared cell extract and subjected it to immunoblotting with specific polyclonal antibodies (prepared by K.T.). For TRAF6 RNA blotting, we incubated RAW 264.7 cells in the presence of RANKL with or without IFN- γ (100 U ml⁻¹). We loaded 5 μ g of total RNA in each lane, and used the 1.7-kilobase (kb) fragment of mouse TRAF6 cDNA (nucleotides 584–2312; ref. 18) as a probe.

Retroviral gene transduction

We constructed a retroviral vector, pMX-TRAF6-IRES-EGFP, by inserting a 2.0-kb EcoRI/StuI fragment of the mouse full-length TRAF6 gene (a gift from J. Inoue) into the same sites in pMX-IRES-EGFP (a gift from T. Kitamura). We performed retrovirus packaging by colipofection of these pMX vectors and pPAMpsi2 (a gift from C. L. Sawyers) into 293T cells. Two days after inoculation, BMMs were assayed for EGFP and TRAF6 expression by flowcytometry and immunoblotting, respectively, and RANKL/M-CSF and IFN- γ (10 U ml⁻¹) were added to BMMs for the osteoclast formation assay. The TRAF6 protein level in BMMs infected with TRAF6 virus was about tenfold more than in control BMMs. We evaluated osteoclastogenesis 3 d after RANKL addition by TRAP staining and bone resorption assay as described³⁰. Most of the TRAP⁺ MNCs are GFP⁺, and more than 50% of the GFP⁺ cells differentiated into TRAP⁺ MNCs. In the absence of RANKL/IFN- γ , the percentage of TRAP⁺ cells/GFP⁺ cells was about 6% in pMX-TRAF6-IRES-EGFP-inoculated BMMs.

Pulse-chase labelling

We inoculated RAW 264.7 cells for 30 min in methionine-free DMEM containing 10% dialysed fetal bovine serum and 100 mCi ml⁻¹ [³⁵S]methionine. We removed the labelling medium and generated whole-cell lysate after incubation for indicated periods in the presence of RANKL with or without IFN- γ (100 U ml⁻¹). We carried out immunoprecipitation with an anti-TRAF6 polyclonal antibody (Santa Cruz) and SDS-polyacrylamide gel electrophoresis, followed by autoradiography. Six hours after stimulation by RANKL alone, about 50% of the pulse-labelled TRAF6 underwent degradation, and addition of IFN- γ resulted in almost complete degradation. We added MG132 (50 μ M, Sigma), lactacystin (20 μ M, Sigma), cycloheximide (20 μ g ml⁻¹, Wako Chemicals), calpain inhibitor II (50 μ M, Sigma) or caspase inhibitor I (50 μ M, Calbiochem) after removal of the labelling medium. We carried out this experiment several times, and obtained essentially the same result.

Ubiquitination of TRAF6

We prepared whole-cell lysate from RAW 264.7 cells, 48 h after the transfection of HA-tagged ubiquitin cDNA. We added RANKL with or without 100 U ml⁻¹ IFN- γ 5 h before protein extraction, and we added 50 μ M MG132 1 h before protein extraction. Immunoprecipitates with an anti-TRAF6 polyclonal antibody (Santa Cruz) was first blotted with an anti-HA antibody (Boehringer). We blotted the same membrane with an anti-TRAF6 monoclonal antibody (Santa Cruz). For the detection of the endogenous ubiquitin in PA28^{+/+} and PA28^{-/-} mice, we performed immunoblotting with an anti-ubiquitin antibody (Novocastra).

Received 12 July; accepted 2 October 2000.

- Roodman, G. D. Cell biology of the osteoclast. *Exp. Hematol.* **27**, 1229–1241 (1999).
- Suda, T. *et al.* Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr. Rev.* **20**, 345–357 (1999).
- Kong, Y. Y. *et al.* OPG is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
- Kong, Y. Y. *et al.* Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. *Nature* **402**, 304–309 (1999).
- Takayanagi, H. *et al.* Suppression of arthritic bone destruction by adenovirus-mediated csk gene transfer to synoviocytes and osteoclasts. *J. Clin. Invest.* **104**, 137–146 (1999).
- Horwood, N. J. *et al.* Activated T lymphocytes support osteoclast formation *in vitro*. *Biochem. Biophys. Res. Commun.* **265**, 144–150 (1999).
- Takahashi, N., Mundy, G. R. & Roodman, G. D. Recombinant human interferon- γ inhibits formation of human osteoclast-like cells. *J. Immunol.* **137**, 3544–3549 (1986).
- Udagawa, N. *et al.* Interleukin-18 (interferon- γ -inducing factor) is produced by osteoblasts and acts via granulocyte/macrophage colony-stimulating factor and not via interferon- γ to inhibit osteoclast formation. *J. Exp. Med.* **185**, 1005–1012 (1997).
- Ukai, T., Hara, Y. & Kato, I. Effects of T cell adoptive transfer into nude mice on alveolar bone resorption induced by endotoxin. *J. Periodontol. Res.* **31**, 414–422 (1996).
- Chiang, C. Y., Kyritsis, G., Graves, D. T. & Amar, S. Interleukin-1 and tumor necrosis factor activities partially account for calvarial bone resorption induced by local injection of lipopolysaccharide. *Infect. Immun.* **67**, 4231–4236 (1999).
- Huang, S. *et al.* Immune response in mice that lack the interferon- γ receptor. *Science* **259**, 1742–1745 (1993).
- Manoury-Schwartz, B. *et al.* High susceptibility to collagen-induced arthritis in mice lacking IFN- γ receptors. *J. Immunol.* **158**, 5501–5506 (1997).
- Vermeire, K. *et al.* Accelerated collagen-induced arthritis in IFN- γ receptor-deficient mice. *J. Immunol.* **158**, 5507–5513 (1997).
- Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. How cells respond to interferons. *Annu. Rev. Biochem.* **67**, 227–264 (1998).
- Taniguchi, T., Lamphier, M. S. & Tanaka, N. IRF-1: the transcription factor linking the interferon response and oncogenesis. *Biochim. Biophys. Acta* **1333**, M9–M17 (1997).
- Wong, B. R. *et al.* The TRAF family of signal transducers mediates NF- κ B activation by the TRANCE receptor. *J. Biol. Chem.* **273**, 28355–28359 (1998).
- Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
- Naito, A. *et al.* Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. *Genes Cells* **4**, 353–362 (1999).
- Baumeister, W., Walz, J., Zuhl, F. & Seemuller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380 (1998).
- Tanaka, K. & Kasahara, M. The MHC class I ligand-generating system: roles of immunoproteasomes and the interferon- γ -inducible proteasome activator PA28. *Immunol. Rev.* **163**, 161–176 (1998).
- Preckel, T. *et al.* Impaired immunoproteasome assembly and immune responses in PA28^{-/-} mice. *Science* **286**, 2162–2165 (1999).
- Tanahashi, N. *et al.* Hybrid proteasomes. Induction by interferon- γ and contribution to ATP-dependent proteolysis. *J. Biol. Chem.* **275**, 14336–14345 (2000).
- Firestein, G. S. & Zvaifler, N. J. How important are T cells in chronic rheumatoid synovitis? *Arthritis Rheum.* **33**, 768–773 (1990).
- Kinne, R. W., Palombo-Kinne, E. & Emrich, F. T-cells in the pathogenesis of rheumatoid arthritis: villains or accomplices? *Biochim. Biophys. Acta* **1360**, 109–141 (1997).
- Takayanagi, H. *et al.* Involvement of receptor activator of nuclear factor κ B ligand/osteoclast differentiation factor in osteoclastogenesis from synoviocytes in rheumatoid arthritis. *Arthritis Rheum.* **43**, 259–269 (2000).
- Gravallese, E. M. *et al.* Synovial tissue in rheumatoid arthritis is a source of osteoclast differentiation factor. *Arthritis Rheum.* **43**, 250–258 (2000).
- Billiau, A. Interferon- γ : biology and role in pathogenesis. *Adv. Immunol.* **62**, 61–130 (1996).
- Matsuyama, T. *et al.* Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell* **75**, 83–97 (1993).
- Meraz, M. A. *et al.* Targeted disruption of the Stat1 gene in mice reveals unexpected physiologic specificity in the JAK-STAT signaling pathway. *Cell* **84**, 431–442 (1996).
- Yasuda, H. *et al.* Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc. Natl Acad. Sci. USA* **95**, 3597–3602 (1998).

Acknowledgements

We thank J. Inoue for the gift of TRAF6 cDNA; T. Kitamura for the pMX vectors; E. Barsoumian for critical reading of the manuscript; I. Kawai for assistance; and M. Asagiri for discussion. This work was supported by a grant for Advanced Research on Cancer from the Ministry of Education, Science, Sports and Culture of Japan; a research fellowship of the Japan Society for the Promotion of Science for Young Scientists; Health Sciences Research Grants from the Ministry of Health and Welfare of Japan; and a grant from Japan Orthopaedics and Traumatology Foundation.

Correspondence and requests for materials should be addressed to T.T. (e-mail: tada@m.u-tokyo.ac.jp).

Development of a preventive vaccine for Ebola virus infection in primates

Nancy J. Sullivan*, Anthony Sanchez†, Pierre E. Rollin†, Zhi-yong Yang* & Gary J. Nabel*

* Vaccine Research Center, National Institutes of Health, 40 Convent Drive, MSC-3005, Bethesda, Maryland 20892, USA

† Special Pathogens Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Mailstop G14, Atlanta, Georgia 30333, USA

Outbreaks of haemorrhagic fever caused by the Ebola virus are associated with high mortality rates that are a distinguishing feature of this human pathogen. The highest lethality is associated with the Zaire subtype, one of four strains identified to date^{1,2}. Its rapid progression allows little opportunity to develop natural immunity, and there is currently no effective anti-viral therapy. Therefore, vaccination offers a promising intervention to prevent infection and limit spread. Here we describe a highly effective vaccine strategy for Ebola virus infection in non-human primates. A combination of DNA immunization and boosting with adenoviral vectors that encode viral proteins generated cellular and humoral immunity in cynomolgus macaques. Challenge with a lethal dose of the highly pathogenic, wild-type, 1976 Mayinga strain of Ebola Zaire virus resulted in uniform infection in controls, who progressed to a moribund state and death in less than one week. In contrast, all vaccinated animals were asymptomatic for more than six months, with no detectable virus after the initial challenge. These findings demonstrate that it is possible to develop a preventive vaccine against Ebola virus infection in primates.

Genetic immunization has been shown to influence both humoral and cellular immune activation pathways and to protect against infection by human pathogens^{3–6}. The effectiveness of plasmid vaccines is thought to result from host-cell protein synthesis and endogenous presentation of the immunogen, and possibly to immunostimulatory effects of plasmid DNA itself^{7,8}. DNA vaccines have been shown to elicit specific immune responses to Ebola virus antigens and to protect guinea-pigs⁹ and mice¹⁰ against challenge with Ebola virus adapted to produce lethal infection in rodents^{11,12}. Although both cell-mediated and humoral immune responses were elicited, antibody titre correlated with the degree of protection in animals immunized with plasmids encoding proteins

Table 1 Multivalent genetic immunization in guinea-pigs

ID	Immunization	ELISA IgG	Survival
1	Plasmid	0	No
2	Plasmid	0	No
3	Plasmid	0	No
4	Plasmid	0	No
5	GP(Z)	6,400	Yes
6	GP(Z)	6,400	Yes
7	GP(Z)	6,400	Yes
8	GP(Z)	3,200	Yes
9	GP(Z) + NP	6,400	Yes
10	GP(Z) + NP	6,400	Yes
11	GP(Z) + NP	6,400	Yes
12	GP(Z) + NP	6,400	Yes
13	GP(Z,IC,S) + NP	6,400	Yes
14	GP(Z,IC,S) + NP	1,600	Yes
15	GP(Z,IC,S) + NP	6,400	Yes
16	GP(Z,IC,S) + NP	6,400	Yes

Guinea-pigs were immunized intramuscularly three times at two-week intervals with 100 μ g of DNA (Plasmid, 100 μ g p1012; GP(Z), 100 μ g pGP(Z); GP(Z) + NP, 75 μ g pGP(Z) and 25 μ g pNP; GP(Z, IC, S) + NP, 25 μ g each of pGP(Z), pGP(IC), pGP(S) and pNP). Serum was collected six weeks after the first injection and pre-challenge titres for antibody to Ebola GP (ELISA IgG) were measured by ELISA²⁸ and are displayed as the reciprocal end-point dilution. Three months after the final immunization the animals were challenged as described previously⁹.

from the Zaire subtype of Ebola virus.

A broadly effective vaccine would need to provide immunity to the multiple Ebola subtypes isolated in human infections (Zaire, Sudan and Ivory Coast), but a multivalent vaccine might dilute the specific immune response demonstrated for the single subtype vaccine. To address this concern, we analysed the efficacy of the original Ebola Zaire DNA vaccine in comparison to its use in combination with DNA from Ebola subtypes Sudan and Ivory Coast. As in a previous study⁹, immunization with a single plasmid encoding Zaire-subtype virion glycoprotein, (GP(Z)), generated a substantial virus-specific antibody response and conferred protective immunity in guinea-pigs (Table 1). Inclusion of a plasmid expressing Ebola nucleoprotein, NP, did not affect the antibody titre to Ebola GP(Z) or diminish its protective efficacy. Further broadening of the vaccine components to include NP and three subtypes of Ebola glycoprotein, Zaire, Ivory Coast and Sudan (GP(Z,IC,S) + NP), yielded a pre-challenge immune response comparable to the single-plasmid vaccine. Moreover, complete protection from infection with Ebola Zaire was achieved in guinea-pigs that received the multivalent vaccine (Table 1, subjects 13–16). Anamnestic antibody was not induced by the virus challenge, indicating that the vaccine itself provided an immune response sufficient to efficiently clear the virus. These findings show that multivalent plasmid immunization did not substantially diminish glycoprotein (GP)-specific antibody production and its protective efficacy in a rodent model.

Because protection in the rodent model of Ebola virus infection correlated with antibody titres, and efficient humoral responses may influence clinical outcome in human disease^{13,14}, we considered it important to elicit a strong humoral response for vaccines tested in primates, although cell-mediated immunity is coordinately induced and probably contributes to protection⁹. Recently, regimens of DNA priming followed by administration of viral vectors have demonstrated enhanced immune responses compared to vaccines using DNA alone^{15–18}. Recombinant, replication-deficient adenoviruses can be grown to high titre, infect antigen-presenting cells, and induce potent immune responses^{19–21}. Adenoviruses have shown a boosting effect in mice²², but the combination of DNA and adenovirus has not been tested for efficacy in an infectious challenge model, and the success of this approach in primates is yet unknown.

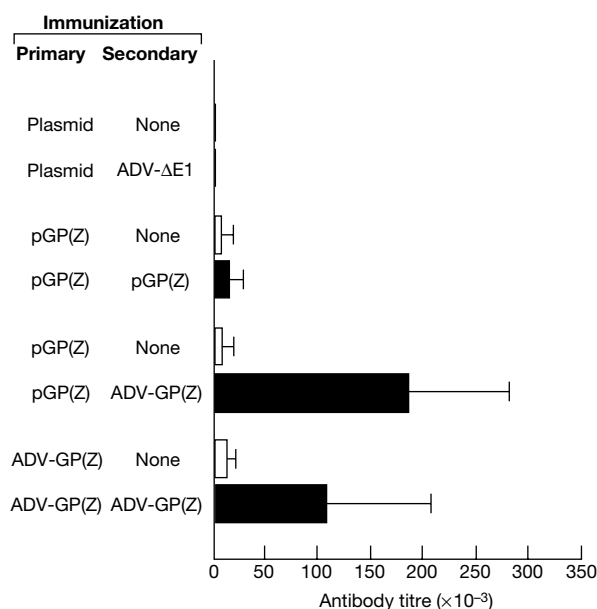


Figure 1 Ebola-specific antibody responses generated by different DNA/adenovirus prime-boost combinations. See Methods. Data are the means of the reciprocal end-point dilution for each group of mice and error bars represent the standard deviation.

We therefore developed a recombinant adenoviral vector that directs high-level GP expression (ADV-GP(Z)) and used this vector to test whether a modified prime-boost strategy would augment the antibody response to Ebola virus obtained with naked DNA alone. Mice were injected with DNA and adenovirus vectors either singly or in combinations, and cell-mediated and humoral immune responses were assessed. A 10-fold to 100-fold increase in antibody titre was found in mice injected with DNA followed by an adenovirus boost, compared with DNA immunization alone (Fig. 1). An increase in cytotoxic T-cell responses was also observed with this combination (data not shown). Immunization with ADV-GP(Z) alone yielded antibody titres that were not

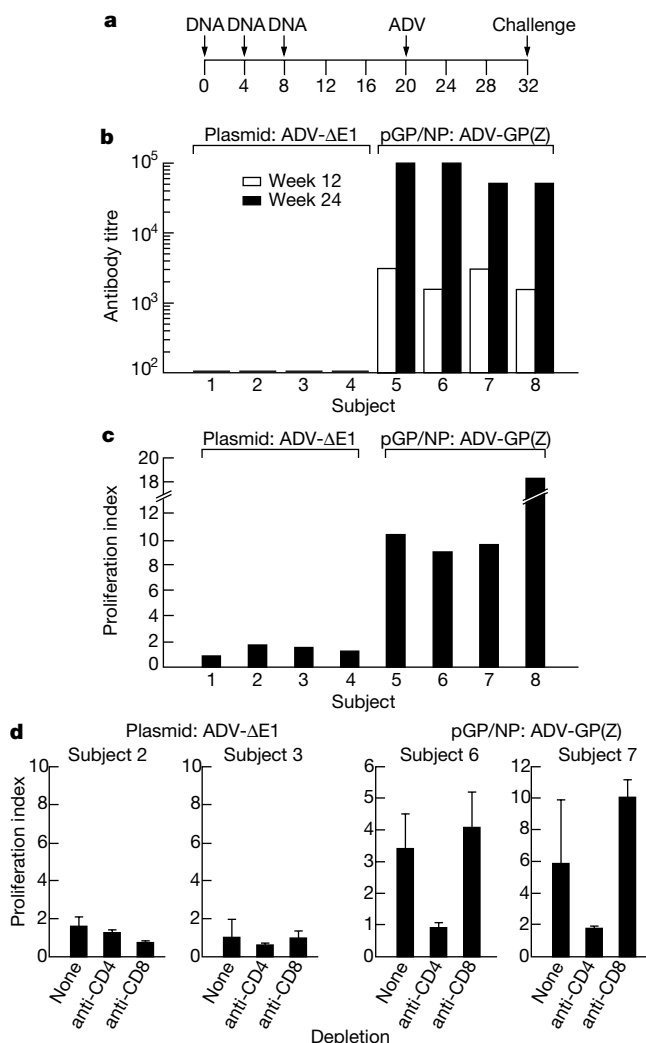


Figure 2 DNA-adenovirus immunization of cynomolgus macaques. **a**, Immunization schedule for DNA and/or adenovirus injections, and challenge with the wild-type Mayinga strain of the Zaire subtype of Ebola virus. **b**, Elisa titres of Ebola-specific antibodies in serum. Serum was collected at week 12 (open bar) and 2 days before immunization at week 24 (closed bar). **c**, Lymphoproliferative responses to Ebola-secreted glycoprotein (sGP) after immunization. Bars represent the average fold-proliferation of all four blood samples for each subject. The standard deviation is not shown because the baseline level of induction varied between experiments. However, PBMC from all eight animals were assayed within the same experiment for each time point, and the averages displayed are representative of the results obtained for any single time point. **d**, Lymphoproliferative responses to Ebola sGP in bulk PBMC following depletion of lymphocyte subsets. PBMC from week 24 were treated with Dynal magnetic beads coated with the indicated antibody to deplete CD4⁺ or CD8⁺ cell subsets. Cells remaining after depletion were normalized for input cell number and stimulated as described in the Methods. Results are shown for two control (subjects 2 and 3) and two vaccinated (subjects 6 and 7) monkeys.

significantly different from those obtained with the DNA prime, adenovirus boost, immunization. These data suggest that immunogenicity of the Ebola GP DNA vaccine in mice is improved by boosting with recombinant adenovirus and that this strategy might represent a useful approach to enhance immune responses in non-human primates.

Whereas the rodent model has been useful in the development of a vaccine strategy, Ebola virus isolated directly from humans must first be adapted by multiple, sequential passages in rodents in order to produce a lethal infection in mice or guinea-pigs^{11,12}. Primate models of Ebola infection are thought to have a stronger predictive value for human disease and immune protection. We therefore

conducted studies in non-human primates using a bimodal DNA/ADV vaccine and the multiple plasmid strategy that correlated with protection in guinea-pigs. Cynomolgus macaques (*Macaca fascicularis*) received three injections of naked DNA vectors at four-week intervals (Fig. 2a) and, after several months of rest, which has been shown to boost immune responses²³, were boosted with recombinant adenovirus expressing only the Zaire glycoprotein (Fig. 2a). Control animals received empty vectors (plasmid DNA and ADV-ΔE1 recombinant adenovirus), and vaccinated animals received the multicomponent DNA vaccine containing NP and three subtypes of Ebola GP (pGP/NP), followed by ADV-GP(Z). As expected, anti-Ebola serum antibodies could not be detected in control animals, but in animals receiving the Ebola vaccine, an antigen-specific antibody response was detected at week 12, one month after the third DNA injection (Fig. 2b). After boosting with recombinant adenovirus, antibody titres increased 10-fold to 20-fold over the levels obtained with DNA alone. Three months after the final immunization, antibody levels remained high (data not shown), except for one animal (subject 8) whose titre dropped slightly from 5×10^4 to 1.3×10^4 .

Primate cellular responses to Ebola antigens were next examined with an *in vitro* lymphocyte-proliferation assay. In control monkeys, antigen-specific lymphocyte proliferation, measured by ³H-thymidine uptake, was equivalent to that in matched, unstimulated cells, resulting in a proliferation index near 1.0 for each animal (Fig. 2c). In contrast, peripheral blood mononuclear cells (PBMC) from animals immunized with the multivalent vaccine showed 9-fold to 20-fold increased stimulation, demonstrating a robust immune response to Ebola antigen at the cellular level. Depletion of CD4-positive lymphocytes reduced the antigen-stimulated proliferative response of PBMC from vaccinated monkeys to the level observed in control animals (Fig. 2d). Depletion of CD8-positive lymphocytes, however, did not affect Ebola-antigen-specific lymphocyte proliferation. Therefore, the CD4-positive subset of lymphocytes, which provide the T-cell help required for high antibody titres, contributes to the vaccine-induced cellular immune response.

To determine the protective efficacy of this vaccination regimen, monkeys were challenged with a lethal dose of the wild-type Mayinga strain from the Zaire subtype of Ebola virus. In the control monkeys, blood chemistry revealed an increase in hepatic enzymes (Fig. 3a and b) that is characteristic for Ebola virus infection²⁴. No such increase was observed in vaccinated subjects. The elevation of serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) was parallel to a dramatic increase in viraemia in all of the control animals (Fig. 3c). In contrast, no substantial increase in viral load was observed in vaccinated monkeys. The kinetics of disease progression was similar among the control animals, and the disease incidence was 100% in this group. Death occurred between days 5 and 6 for 3 animals, and the last monkey, moribund, was euthanized on day 7 (Fig. 3c). In contrast, 4 out of 4 monkeys immunized with the combination DNA-adenovirus vaccine survived this lethal challenge of Ebola virus, and sterilizing immunity was achieved in 3 out of 4 subjects. The remaining animal showed a small transient rise in viral antigen; however, when followed long-term, all vaccinated animals showed no signs or symptoms of infection, and there was no detectable viraemia for more than six months after infection, as measured by ELISA detection of viral antigen (Fig. 3a) and end-point titration analysis of cultured virus (data not shown). The vaccine recipient (subject 8) that exhibited a transient low level of viraemia on day 10 had undetectable levels by day 17.

As the natural reservoir for Ebola virus is unknown, the potential for traditional public health measures to prevent future outbreaks is limited, thus increasing the urgency for the development of a vaccine and therapeutics in humans. The present findings demonstrate that primates can be immunized against the lethal effects of Ebola virus infection, and that sterilizing immunity can be achieved

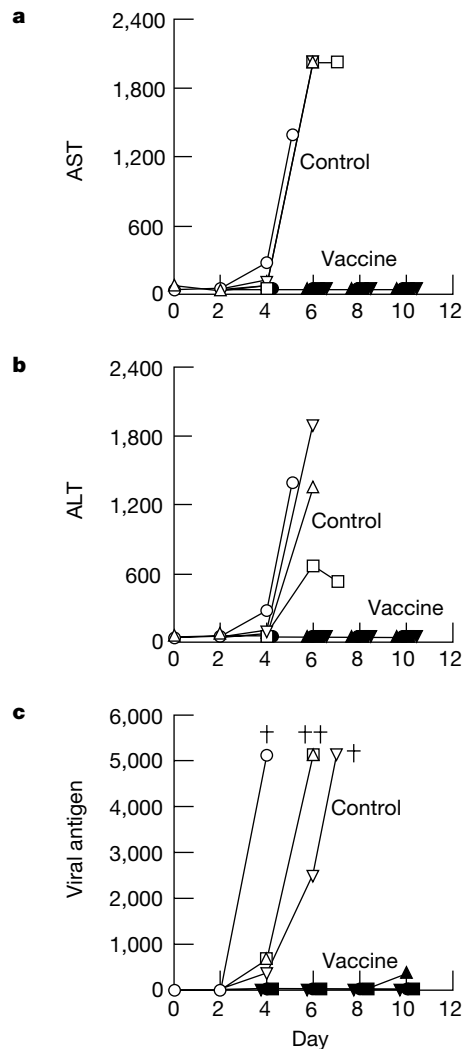


Figure 3 Protection of cynomolgus macaques against lethal challenge with Ebola virus after DNA-adenovirus immunization. **a, b**, Hepatic enzyme levels in monkeys after challenge with Ebola virus. Liver enzymes (alanine aminotransferase (ALT) and aspartate aminotransferase (AST)) levels in the non-human primate sera were measured by standard recommended procedures using a General chemistry 12 reagent disk for the Piccolo Analyzer (Abaxis). Results are shown for four immunized (closed symbols) and four control (open symbols) monkeys. **c**, Plasma viraemia in monkeys after infection with Ebola virus. Crosses represent time of death in control animals (days 5 (subject 1) and 6 (subjects 2 and 4)). One control animal, subject 3, was euthanized on day 7 when it was moribund (see Methods). One vaccinated animal that was resistant to infection, subject 5, was euthanized on day 10 for histological examination of tissues. By day 17, none of the animals had detectable viraemia, and they remained aviremic for the duration of the observation period (six months). Data are the reciprocal end-point dilution of serum for each monkey. Results are shown for four immunized (closed symbols) and four control (open symbols) monkeys.

using a heterologous prime-boost immunization strategy. A multi-component genetic vaccine expressing Ebola virus structural proteins from diverse geographic isolates generated a strong antigen-specific immune response and resulted in the survival of all immunized primates after challenge with a lethal dose of Ebola Zaire, the subtype of this virus associated with the highest number of deaths in human infections. The results of this study suggest that T-cell mediated and humoral immunity contribute to virus clearance in non-human primates, consistent with previous studies in rodents^{9,25}. Two immune parameters, antibody titre (1:75,000 versus <1:100, $P = 0.001$) and the cellular proliferative response (~12-fold versus 1.4-fold, $P = 0.0014$), provided highly significant immune correlates of protection. Studies investigating the correlates of immune protection from Ebola virus infection in humans are hampered by the aggressive nature of the virus and necessarily high level of biosafety containment. With the model of primate immunity presented here, it should now be possible to elucidate the mechanisms of immune protection from Ebola virus infection, to advance immune-based anti-viral therapies, and to develop a human vaccine for this pathogen and possibly other infectious causes of haemorrhagic fever. □

Methods

Vector construction

The construction of DNA vectors expressing Ebola Zaire GP, secreted GP (sGP), and NP has been described previously⁹. The GP Sudan (Genbank accession number U28134) and Ivory Coast (Genbank accession number U28066) expression vectors were constructed similarly. Briefly, GP open reading frames were generated from polymerase chain reaction after reverse transcription of RNA (RT-PCR) products of infected cell RNA using the following primers: 5'-ATC TTC AGG ATC TCG CCA TGG A-3' (Sudan GP gene; NcoI > ATG), 5'-GAT ATT CAA CAA AGC AGC TTG CAG-3' (Sudan GP gene; C-terminus GP stop), 5'-CTA ATC ACA GTC ACC ATG GGA-3' (IC GP gene; NcoI > ATG), 5'-AAA GTA TGC TAT ATT AGT TCA-3' (Ivory Coast GP gene; carboxy-terminus GP stop) yielding the TA clones PCR2.1 Sudan and PCR2.1 Ivory Coast. The Sudan glycoprotein was digested from plasmid PCR2.1 with XbaI/HindIII, Klenow treated, and cloned into the XbaI site of p1012 (ref. 9). Ivory Coast GP was digested from plasmid PCR2.1 with EcoRI, Klenow treated, and cloned into the XbaI site of p1012 (ref. 9).

To make ADV-GP, the BamHI/EcoRI fragment of GP(Z) was digested from PGEM-3Zf(-)-GP, treated with Klenow, and inserted into HindIII/XbaI/Kle/CIP-treated pRC/CMV plasmid. The resulting plasmid (PRC/CMV-GP(Z)) was digested by NruI/DraIII and treated with Klenow. The NruI/DraIII/Kle fragment containing the CMV enhancer, GP(Z) DNA and bovine growth hormone polyadenylation signal was inserted into the BglII site of the adenoviral shuttle plasmid pAdBgIII²⁶. The adenovirus, a first generation dl 309-based Ad5 vector, contained a deletion in E1 to render the vector replication defective and a partial deletion/substitution in E3, which disrupts the coding sequences for the E3 proteins with a relative molecular mass of 14,700, 14,500 and 10,400 (M_r 14.7, 14.5 and 10.4 K), respectively. The recombinant adenovirus expressing Zaire GP, ADV-GP(Z), was made according to previously published methods²⁷. The dose of adenovirus administered, 10¹⁰ plaque-forming units (PFU) per animal (approximately 3 × 10 PFU kg⁻¹), is within the range used safely in human gene-therapy trials.

Animal study and safety

Eight cynomolgus macaques (*M. fascicularis*), 3 years old and weighing 2–3 kg, obtained from Covance, were used for the immunization and challenge experiment. To obtain blood specimens and administer vaccines, the monkeys were anaesthetized with Ketamine. The animals were housed singly and received regular enrichment according to the Guide for the Care and Use of Laboratory Animals (DHEW No. NIH 86–23). Just before the Ebola virus challenge and up to the end of the experiment, the animals were maintained in the Maximum Containment Laboratory (BSL-4) and fed and checked daily. One animal was euthanized that appeared moribund and was subsequently necropsied for pathologic examination. In addition, a single asymptomatic vaccinated animal was euthanized for pathologic and virologic analysis.

Mouse immunization

DNA and adenovirus vectors expressing Ebola Zaire glycoprotein or nucleoprotein (GP or NP) were constructed as described previously^{9,26} with gene expression under the control of the cytomegalovirus enhancer and promoter. Mice were immunized intramuscularly with 100 µg of DNA (pGP or a p1012 plasmid control) or 10⁸ PFU of adenovirus (ADV-GP or ADV-ΔE1 control virus) on days 0, 14 and 28, and blood was collected on day 28 (open bars). On day 42, mice received an intramuscular boost with DNA or adenovirus and titres were re-measured on day 56 (solid bars). ELISA IgG titres were determined using 96-well plates coated with a preparation of Ebola virus antigen derived from purified virions and enriched for membrane-associated proteins (GP, VP40 and VP24)²⁸. Specific antigen binding was detected using a goat anti-human IgG(H+L)-horseradish peroxidase conjugate and ABTS/peroxide (substrate/indicator).

Macaque immunization

For the DNA immunizations, animals received 1 mg each of DNA expressing GP(Zaire), [GP(Z)], GP(Ivory Coast) [pGP(IC)], GP(Sudan) [pGP(S)] and NP(Zaire) administered as a mixture [pGP/NP], or 4 mg empty [pGP(Z)] control plasmid bilaterally (2 mg per side) in the deltoid muscle. Immunization at weeks 0 and 4 were by IM injection, and at week 8 by Biojector. For the adenovirus boost, animals received 10¹⁰ PFU of ADV-GP (Zaire subtype) or ADV-ΔE1 (empty vector) divided into two doses administered bilaterally in the deltoid muscle. At week 32, all animals received an intraperitoneal injection of approximately six PFUs of Ebola virus (Zaire 1976 isolate; Mayinga strain)²⁹ in 1 ml Hanks' buffered salt solution. The virus was isolated directly from patient blood and used after a single passage in Vero cells.

ELISA IgG titres were determined as above for control (Plasmid: ADV-ΔE1) and immunized (pGP/NP: ADV-GP(Z)) monkeys. The reciprocal end-point of dilution for each subject was at week 12 and week 24. Serum antibody levels were measured by ELISA as described²⁸. Blood was collected from control (plasmid: ADV-ΔE1) or immunized (pGP/NP: ADV-GP(Z)) animals 1–3 days before immunizations at weeks 4, 8 and 20, and at week 24. Blood was separated over a Percoll gradient to obtain the lymphocyte enriched population. Lymphocytes were stimulated as described previously⁹ for five days *in vitro* using supernatant from cells transfected with either sGP or empty plasmid, and proliferation was measured by ³H-thymidine uptake. The proliferation index was calculated as the proliferation in wells receiving sGP divided by proliferation in wells receiving control supernatant.

Viral detection in macaques

The presence of circulating Ebola virus antigen was detected as described²⁸ by capturing VP40 protein from serial dilutions of monkey plasma. 96-well plates coated with anti-VP40 mAb were used to capture antigen, and detection was with a rabbit anti-Ebola virus serum.

Received 4 September; accepted 20 October 2000.

- Feldmann, H., Nichol, S. T., Klenk, H. D., Peters, C. J. & Sanchez, A. Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein. *Virology* **199**, 469–473 (1994).
- Sanchez, A., Trappier, S. G., Mahy, B. W. J., Peters, C. J. & Nichol, S. T. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl Acad. Sci. USA* **93**, 3602–3607 (1996).
- Tang, D. C., DeVit, M. & Johnston, S. A. Genetic immunization is a simple method for eliciting an immune response. *Nature* **356**, 152–154 (1992).
- Ulmer, J. B. *et al.* Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science* **259**, 1745–1749 (1993).
- Wang, B. *et al.* Gene inoculation generates immune responses against human immunodeficiency virus type 1. *Proc. Natl Acad. Sci. USA* **90**, 4156–4160 (1993).
- Sedegah, M., Hedstrom, R., Hobart, P. & Hoffman, S. L. Protection against malaria by immunization with plasmid DNA encoding circumsporozoite protein. *Proc. Natl Acad. Sci. USA* **91**, 9866–9870 (1994).
- Krieg, A. M. *et al.* CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* **374**, 546–549 (1995).
- Sato, Y. *et al.* Immunostimulatory DNA sequences necessary for effective intradermal gene immunization. *Science* **273**, 352–354 (1996).
- Xu, L. *et al.* Immunization for Ebola virus infection. *Nature Med.* **4**, 37–42 (1998).
- Vanderzanden, L. *et al.* DNA vaccines expressing either the GP or NP genes of Ebola virus protect mice from lethal challenge. *Virology* **246**, 134–144 (1998).
- Connolly, B. M. *et al.* Pathogenesis of experimental Ebola virus infection in guinea pigs. *J. Infect. Dis.* **179**, S203–S217 (1999).
- Bray, M., Davis, K., Geisbert, T., Schmaljohn, C. & Huggins, J. A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. *J. Infect. Dis.* **178**, 651–661 (1998).
- Baize, S. *et al.* Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. *Nature Med.* **5**, 423–426 (1999).
- Maruyama, T. *et al.* Ebola virus can be effectively neutralized by antibody produced in natural human infection. *J. Virol.* **73**, 6024–6030 (1999).
- Sedegah, M. *et al.* Boosting with recombinant vaccinia increases immunogenicity and protective efficacy of malaria DNA vaccine. *Proc. Natl Acad. Sci. USA* **95**, 7648–7653 (1998).
- Hanke, T. *et al.* Enhancement of MHC class I-restricted peptide-specific T cell induction by a DNA prime/MVA boost vaccination regime. *Vaccine* **16**, 439–445 (1998).
- Robinson, H. L. *et al.* Neutralizing antibody-independent containment of immunodeficiency virus challenges by DNA priming and recombinant pox virus booster immunizations. *Nature Med.* **5**, 526–534 (1999).
- Schneider, J. *et al.* Enhanced immunogenicity for CD8+ T cell induction and complete protective efficacy of malaria DNA vaccination by boosting with modified vaccinia virus Ankara. *Nature Med.* **4**, 397–402 (1998).
- Davis, A. R. *et al.* Expression of hepatitis B surface antigen with a recombinant adenovirus. *Proc. Natl Acad. Sci. USA* **82**, 7560–7564 (1985).
- Natuk, R. J. *et al.* Adenovirus-human immunodeficiency virus (HIV) envelope recombinant vaccines elicit high-titered HIV-neutralizing antibodies in the dog model. *Proc. Natl Acad. Sci. USA* **89**, 7777–7781 (1992).
- Xiang, Z. Q., Yang, Y., Wilson, J. M. & Ertl, H. C. A replication-defective human adenovirus recombinant serves as a highly efficacious vaccine carrier. *Virology* **219**, 220–227 (1996).
- Xiang, Z. Q., Pasquini, S. & Ertl, H. C. Induction of genital immunity by DNA priming and intranasal booster immunization with a replication-defective adenoviral recombinant. *J. Immunol.* **162**, 6716–6723 (1999).
- Letvin, N. L. *et al.* Potent, protective anti-HIV immune responses generated by bimodal HIV envelope DNA plus protein vaccination. *Proc. Natl Acad. Sci. USA* **94**, 9378–9383 (1997).
- Fisher-Hoch, S. P. *et al.* Pathophysiology of shock and hemorrhage in a fulminating viral infection (Ebola). *J. Infect. Dis.* **152**, 887–894 (1985).

25. Wilson, J. *et al.* Epitopes involved in antibody-mediated protection from Ebola virus. *Science* **287**, 1664–1666 (2000).
26. Ohno, T. *et al.* Gene therapy for vascular smooth muscle cell proliferation after arterial injury. *Science* **265**, 781–784 (1994).
27. Aoki, K., Barker, C., Danthine, X., Imperiale, M. J. & Nabel, G. J. Efficient generation of recombinant adenoviral vectors by Cre-lox recombination *in vitro*. *Mol. Med.* **5**, 224–231 (1999).
28. Ksiazek, T. G. *et al.* Enzyme immunoassay for Ebola virus antigens in tissues of infected primates. *J. Clin. Microbiol.* **30**, 947–950 (1992).
29. Kiley, M. P., Regnery, R. L. & Johnson, K. M. Ebola virus: identification of virion structural proteins. *J. Gen. Virol.* **49**, 333–341 (1980).

Acknowledgements

We thank J. DeCoster and R. Grecoe for technical assistance with primates and G. Reynolds for the animal handling in the Biosafety Laboratory-4, K. Reiman and J. Stein for advice, assistance and comments, and N. Barrett, C. Davis and A. Tislerics for help in preparing the manuscript.

Correspondence and requests for materials should be addressed to G.J.N. (e-mail: gnabel@nih.gov).

Coenzyme Q is an obligatory cofactor for uncoupling protein function

Karim S. Echtay*, Edith Winkler* & Martin Klingenberg

Institute of Physiological Chemistry, University of Munich, Schillerstrasse 44, D-80336 Munich, Germany

* These authors contributed equally to this work

Uncoupling proteins (UCPs) are thought to be intricately controlled uncouplers^{1–3} that are responsible for the futile dissipation of mitochondrial chemiosmotic gradients, producing heat rather than ATP. They occur in many animal and plant cells^{4–9} and form a subfamily of the mitochondrial carrier family¹⁰. Physiological uncoupling of oxidative phosphorylation must be strongly regulated to avoid deterioration of the energy supply and cell death, which is caused by toxic uncouplers. However, an H⁺ transporting uncoupling function is well established only for UCP1 from brown adipose tissue^{2,8,9,11}, and the regulation of UCP1 by fatty acids, nucleotides and pH remains controversial^{2,12–14}. The failure of UCP1 expressed in *Escherichia coli* inclusion bodies to carry out fatty-acid-dependent H⁺ transport activity inclusion bodies¹⁵ made us seek a native UCP cofactor. Here we report the identification of coenzyme Q (ubiquinone) as such a cofactor. On addition of CoQ₁₀ to reconstituted UCP1 from inclusion bodies, fatty-acid-dependent H⁺ transport reached the same rate as with native UCP1. The H⁺ transport was highly sensitive to purine nucleotides, and activated only by oxidized but not reduced CoQ. H⁺ transport of native UCP1 correlated with the endogenous CoQ content.

Whereas UCP1 from brown adipose tissue (BAT) can be isolated and is thus functionally well characterized¹¹, the isolation of more recently discovered UCPs has not been possible because of their low tissue content^{8,9}. For this reason and for mutagenesis studies, we expressed recombinant UCPs in yeast and in *E. coli*^{15–20}. Our inability to reconstitute H⁺ transport activity from UCP1 expressed as inclusion bodies (IB-UCP1) in *E. coli* prompted us to consider several causes. (1) The reconstituted IB-UCP1 may not attain the full native configuration. (2) The reconstituted IB-UCP1 may still contain traces of the anion detergent sarcosyl used to solubilize UCP1, which, as a fatty-acid analogue, may prevent fatty-acid binding. (3) A covalent or (4) non-covalent cofactor present in the native UCP1 may be missing.

To first bring soluble IB-UCP1 into a native state, we thoroughly removed the sarcosyl used for solubilization of inclusion bodies by

substitution with digitonin and anion-exchange treatment. We succeeded in restoring the nucleotide-binding capacity of soluble IB-UCP1 (Fig. 1a) with the fluorescent derivative dansyl-GTP, which had been used to test the nucleotide binding of native UCP1 (ref. 21). The decrease of fluorescence caused by competition with excess ATP corresponded to the specific binding of UCP1. We further tested for native protein configuration with the carboxyl reagent, Woodward reagent K, which blocks with high specificity

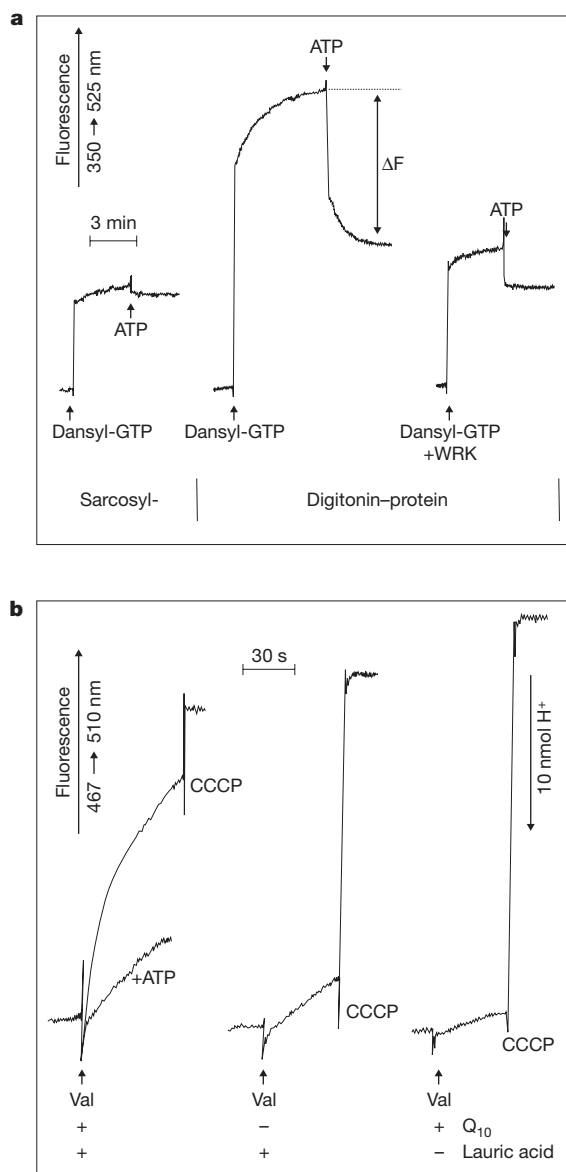


Figure 1 Functional integrity of refolded recombinant IB-UCP1 from *E. coli*.

a, Fluorescence response of 2'-O-(dimethylamino-naphthalene-1-sulfonyl)-GTP (dansyl-GTP) binding to sarcosyl- and digitonin-treated IB-UCP1. Dansyl-GTP (5 μM) was added to a solution of 45 μg ml⁻¹ UCP1 in 10 mM MES/HEPES buffer containing 0.3% digitonin, pH 6.5 at 10 °C. To differentiate the UCP1-linked binding, 0.5 mM ATP was added to displace the fluorescent ligand. To confirm binding to native UCP1, 10 μM Woodward reagent K was added 30 min before dansyl-GTP. **b**, Recording of H⁺ influx into vesicles reconstituted with digitonin-treated IB-UCP1. H⁺ influx was monitored by the fluorescence of pyranine. Vesicles (50 μl), containing 1.25 μg protein and 0.42 mg phospholipid, were added to a medium containing 0.5 mM HEPES pH 7.3, 1 μM pyranine, 0.5 mM EDTA, 280 mM sucrose in a final volume of 330 μl at 10 °C; 125 μM lauric acid and 2 nmol CoQ₁₀ were added where indicated. Valinomycin (Val; 2.5 μM) was added to induce K⁺ efflux and thus a membrane potential. The uncoupler carbonyl cyanide m-chlorophenylhydrazone (CCCP) (1 μM) was added to determine the vesicle capacity for H⁺ uptake.

the carboxyl group E190, necessary for nucleotide binding^{18,22}. On incorporation into vesicles, UCP1 was able to transport Cl^- but not H^+ . As these results excluded causes (1) and (2) above, it seemed most probable that a mitochondrial cofactor was missing from IB-UCP1. The lack of a covalent component (3) could also be excluded, because after partial removal of the detergent Triton X-100, mitochondrial UCP1 could be brought in a state where it was also unable to transport H^+ , although Cl^- transport and inhibition by nucleotides were restored (data not shown). The existence of a noncovalent cofactor remained the most logical explanation.

On the assumption that it was a lipid cofactor, we extracted lyophilized mitochondria using lipid solvents and screened the extracts for their ability to activate H^+ transport. Chloroform extracts were prepared not only from mitochondria of BAT and

UCP1-expressing yeast, but also from UCP1-free bovine heart, to determine whether the activating component is confined to UCP1-containing mitochondria. The rates shown in Fig. 2a were evaluated from the recordings of H^+ uptake (Fig. 1b). All extracts, independent of whether mitochondria contained UCP1, stimulated H^+

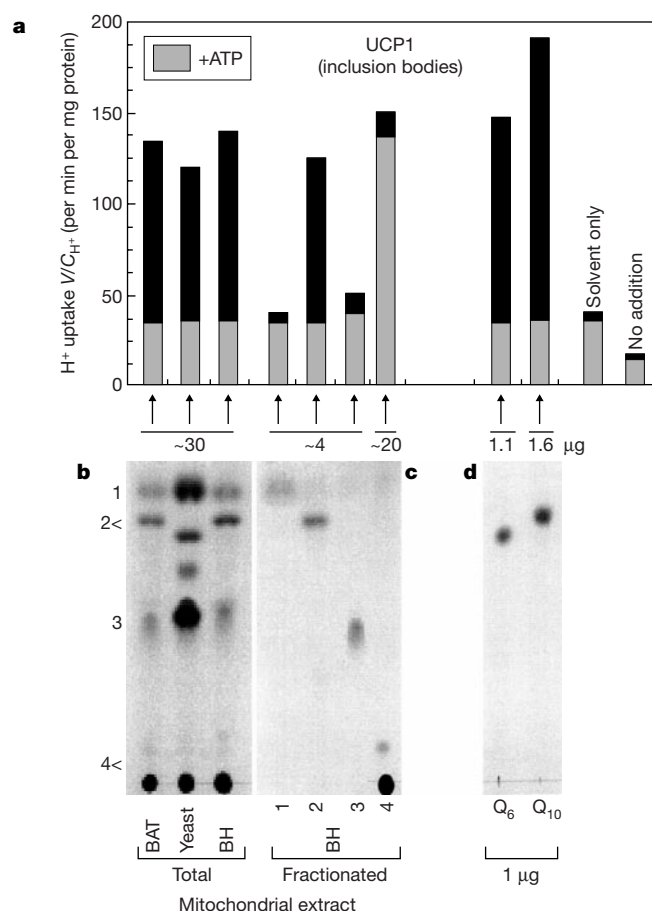


Figure 2 Identification of CoQ as the activating component in mitochondrial lipids of H^+ transport by UCP1. **a**, The activation of H^+ transport in reconstituted digitonin-treated UCP1 proteoliposomes by total lipid extracts from mitochondria (BAT, brown adipose tissue; BH, bovine heart), by fractionated lipids from BH and by CoQ₆ and CoQ₁₀, and in the presence of 125 μM lauric acid. Black bar corresponds to the H^+ transport portion sensitive to 20 μM ATP. Added lipids are analysed in the silica thin-layer chromatograms. H^+ transport rates are evaluated from recordings of H^+ uptake, as shown in Fig. 1. Results are presented as initial H^+ transport rates (V , μmol per min per mg protein) divided by the capacity of the vesicles (C , μmol) determined by addition of CCCP. The transport rates range from 30 to 120 μmol per min per mg protein. **b**, Mitochondrial extracts on silica thin-layer chromatograms using cyclohexane:ethylacetate (30:20) as the mobile phase. Mitochondria from BAT, yeast and BH were lyophilized and extracted with chloroform overnight at 4 °C. After filtration, the extracts were dried under vacuum; the yield is about 20% lipid by mass of mitochondrial protein. **c**, The chloroform extract (1.5 ml) of BH mitochondria was applied to a silica column (volume 20–25 ml) at room temperature using chloroform for equilibration and elution. Fraction 1, triglycerides; 2, activating compounds; 3, fatty acids; 4, phospholipids. **d**, Thin-layer chromatography of CoQ₆ and CoQ₁₀ (1 μg) under the same conditions.

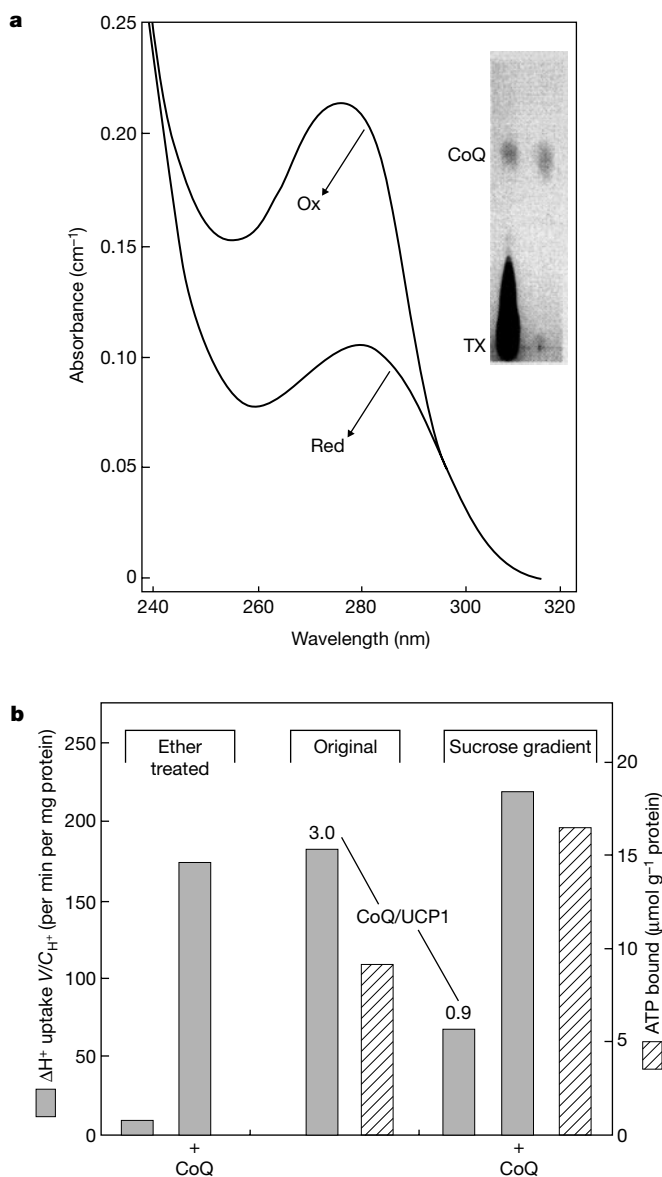


Figure 3 Identification of CoQ in native isolated mitochondrial UCP1. **a**, Absorbance spectrum of CoQ extracted from a native UCP1 preparation. UCP1 (0.5 mg) solubilized from BAT mitochondria by Triton X-100 and purified by hydroxyapatite chromatography was lyophilized and extracted by stirring with 5 ml chloroform overnight at 4 °C. To remove Triton X-100, the extract was applied to a silica column (volume 3 ml, equilibrated with chloroform) (inset). Fractions corresponding to CoQ were pooled and dissolved in 150 μl dichloromethane. CoQ was identified by a decrease in absorbance at 275 nm on addition of NaBH_4 (10 mg ml^{-1}). **b**, Endogenous CoQ is a cofactor of mitochondrial UCP1. Treatment with diethylether for removal of all endogenous CoQ by precipitation of UCP1. UCP1 (2 mg in 1.2 ml of medium) was exposed first to 15 ml and then to 5 ml of ether at pH 5 for 20 h. After centrifugation at 15,000g, the pellet was solubilized and reconstituted as described for inclusion bodies (see Methods). 'Original' corresponds to mitochondrial UCP1 purified by hydroxyapatite. For partial depletion of CoQ, soluble native UCP1 was purified by sucrose gradient¹¹. UCP1 (2.3 mg) was applied to a 12/20% sucrose step gradient and centrifuged for 18 h at 90,000g. The peak protein fractions were analysed for CoQ content. Reconstitution of native UCP1 and ATP binding was determined as described¹⁸.

transport when added to the proteoliposomes. This increase was fully inhibited by 20 μM ATP and thus fulfilled our criterion of caused by UCP1.

We initially separated the activating fraction from the bulk lipids by silica thin-layer chromatography. Good separation was achieved with a mobile phase in which the migration of apolar lipids was favoured and phospholipids stayed at the bottom (Fig. 2b). The lipid patterns were similar from mitochondria both with and

without UCP1 (compare BAT and bovine heart). We extracted the lipid fractions and screened for activator activity as shown in Fig. 2c. Fraction 4, which contained mostly phospholipids, stimulated only ATP-insensitive H^+ transport. Fractions 3 and 1, which contained free fatty acid and triglycerides did not markedly activate H^+ transport. Fraction 2 produced strong activation, which was largely ATP sensitive. Notably, fraction 2 from yeast mitochondria also contained the activator, although it migrated with apparently lower hydrophobicity.

We collected a larger quantity of fraction 2 from bovine heart by silica-gel column chromatography; NMR analysis showed that it contained mostly triglycerides. In an ultraviolet spectrum, however, there was a distinct absorption maximum at 274 nm. On reduction with NaBH_4 , the maximum absorption was reduced and shifted to 290 nm, characteristic of coenzyme Q (CoQ). To identify whether the lipid or CoQ was the activator, we added pure CoQ_{10} to the vesicles and observed a strong stimulation of H^+ transport that could be inhibited by a low concentration of ATP. As shown in the recordings of H^+ uptake (Fig. 1b), CoQ activated H^+ transport only in the presence of fatty acid.

We then determined whether CoQ was only a suitable 'in vitro' activator for the reconstitution of IB-UCP1, or CoQ was a constituent of the active, isolated mitochondrial UCP1. Mitochondrial UCP1 isolated with Triton X-100 was lyophilized and extracted with chloroform. The extract was fractionated on a silica column so that the bulk of lipids and the excess of Triton X-100 were removed (see insert Fig. 3a), which would block the ultraviolet absorption of CoQ. The eluate equivalent to bovine heart fraction 2 of the mitochondrial extract had a well-defined absorption spectrum corresponding to CoQ (Fig. 3). From the absorbance, we calculated a molar ratio for CoQ:UCP of about 3. As the UCP1 preparation contained a considerable amount of phospholipid (a weight ratio of UCP1 to protein of 5), CoQ may have been distributed between the protein and the phospholipid. We partially removed Triton X-100 and phospholipid by the mild diethylether precipitation of UCP1. H^+ transport of the resolubilized and reconstituted UCP1 was fully dependent on CoQ_{10} addition, as with IB-UCP1 (Fig. 3b).

To prove further the physiological role of CoQ, we determined the dependence of mitochondrial UCP1 on CoQ under conditions in which UCP1 retains its native state throughout the soluble phase. This would eliminate any possible renaturing influence of CoQ on UCP1. We used a sucrose gradient centrifugation to decrease the content of CoQ by removing excess detergent/phospholipid micelles¹¹. The amount of UCP1 was determined by ATP binding. The molar ratio CoQ:UCP, as measure by ATP binding, decreased from 3 to 0.9; the H^+ transport activity was reduced to 40%. But by addition of CoQ, the activity reached 120% of the original value (Fig. 3b).

Although other members of the mitochondrial carrier family do not require CoQ for the reconstitution of the recombinant proteins, we considered whether the marginal fatty-acid-stimulated H^+ transport of the ADP/ATP carrier^{23,24} is also stimulated by CoQ. In reconstituted vesicles with recombinant ADP/ATP carrier expressed in *E. coli*, no effect of CoQ was observed on H^+ transport, indicating the specificity of UCP in its requirement for CoQ (data not shown).

Titration of the H^+ transport activation with increasing amounts of CoQ determined saturation to be ~ 2 nMol CoQ (Fig. 4a). Because most CoQ should be absorbed by the phospholipid, the molar ratio of CoQ to phospholipids is critical, and was 1:300 here. The corresponding molar ratio CoQ:UCP is 80:1, indicating a low affinity of UCP1 to CoQ. In mitochondria, the CoQ content corresponds approximately to a density of 1:100 per phospholipid molecule. The dependence on fatty-acid 'concentration' in the presence of saturating amounts of CoQ was nonlinear (Fig. 4b), differing from the linear dependence with reconstituted native UCP1 (ref. 25). Presumably, some fatty acids were sequestered before they became available to the vesicles by the solvent in

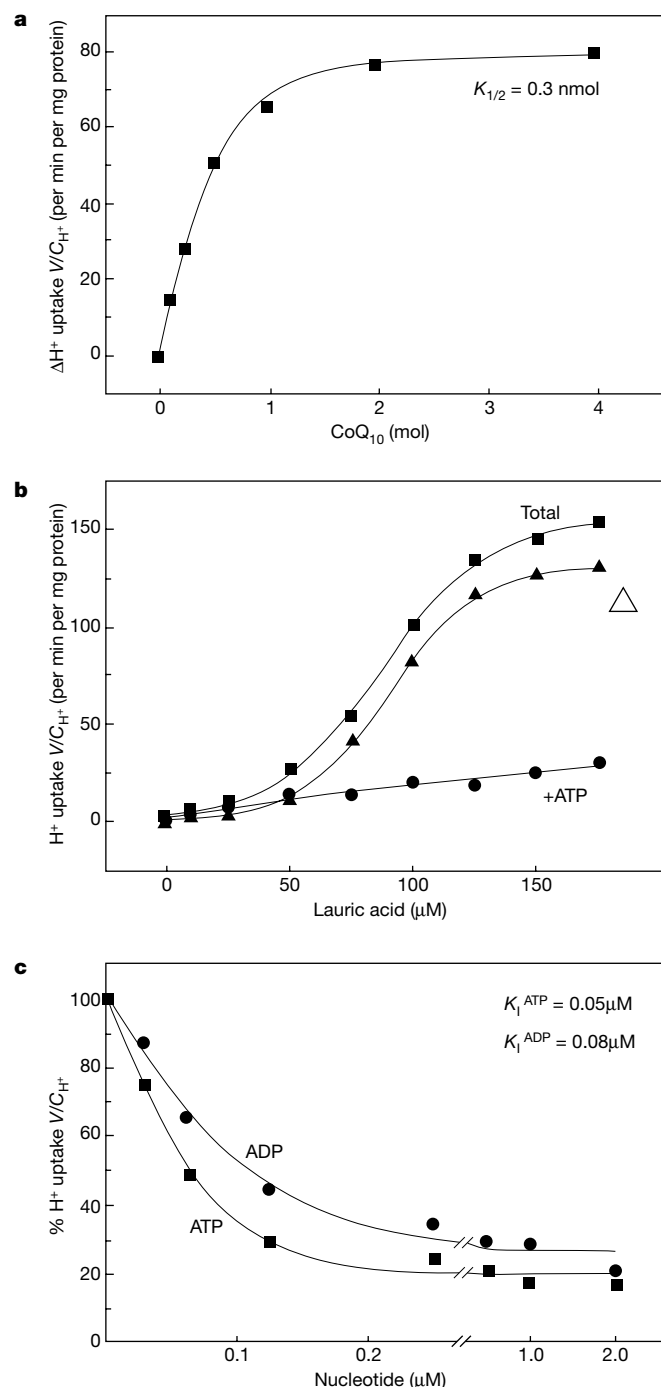


Figure 4 The influence of three regulatory parameters on the H^+ transport activity of recombinant UCP1. **a**, Titration of activity with CoQ_{10} (nmol CoQ_{10} per 0.4 mg phospholipid). H^+ transport by reconstituted IB-UCP1 was measured as shown in Fig. 1, with increasing amount of CoQ_{10} at 125 μM lauric acid. The net transport activity shown was obtained by subtracting the activity in the presence of 20 μM ATP. **b**, Titration with lauric acid at 2 nmol CoQ_{10} . **c**, Dependence of the inhibition on the concentration of ATP and ADP at 125 μM lauric acid and 2 nmol CoQ_{10} .

Table 1 The dependence of H⁺ transport of reconstituted UCP1 on the isoprenoid chain length and on the oxidized and reduced state of CoQ

	UCP1 (<i>E. coli</i>)			UCP1 (BAT) (ether treated)		
	CoQ _{ox}	CoQ _{red}	Q _{ox} + Q _{red}	CoQ _{ox}	CoQ _{red}	CoQ _{ox} + CoQ _{red}
CoQ ₁₀	100	7	88	100	6	85
CoQ ₈	77	7	70	75	8	68
CoQ ₂	25	7	—	30	5	—
CoQ ₁	10	5	—	10	4	—
CoQ ₀	10	7	—	10	6	—

The percentage of activated H⁺ transport rate is shown for reconstituted IB-UCP1 from *E. coli* and for ether-precipitated native UCP1 from BAT, measured as described in Fig. 1 in the presence of 2 nmol CoQ and 125 μM lauric acid. For the competition, 2 nmol CoQ_{ox} + 2 nmol CoQ_{red} were added. Native UCP1 was precipitated with ether, redissolved and treated as described for IB-UCP1. CoQ was reduced by 10 mg ml⁻¹ NaBH₄ and identified by a decrease in absorbance at 275 nm. The maximum transport rate obtained with CoQ₁₀ is equal to 100%. The rates are obtained after subtraction of the activity in the presence of 20 μM ATP.

which CoQ was added. For maximum activity, the concentration of 125 μM lauric acid was the same as with native UCP1. Notably, inhibition of the activated H⁺ transport was highly sensitive to purine nucleotide (Fig. 4c). The maximum inhibition was reached already by 0.25 μM ADP or ATP. The inhibition constant, K_i, for ATP (0.05 μM) and for ADP (0.08 μM) was even lower than those determined with reconstituted native UCP1 (ref. 15), but the ratios K_i^{ATP}:K_i^{ADP} were about the same. The 20% nucleotide insensitive H⁺ transport activity could be attributed to UCP1 molecules that were inverted in the vesicle with the nucleotide-binding site inaccessible from the outside.

We investigated the structural requirements of CoQ for activation by changing the hydrophobic side chain (Table 1). The activation decreased markedly with shorter isoprenoid chains. Only minor activation remained when one or no isoprenoid unit was present (CoQ₁ and CoQ₀). We considered whether or not the activation depended on the redox state of CoQ. The reduced forms of all CoQ homologues were scarcely able to activate H⁺ transport (Table 1), suggesting a regulation of uncoupling *in vivo* by the CoQ_{ox}:CoQ_{red} ratio. There was a possibility that CoQ might undergo a reduction when catalysing H⁺ transport, because, owing to the limited capacity of the vesicles, only five H⁺ ions per CoQ are transported. In the extracts from rapidly frozen and lyophilized vesicles, however, there was no reduced CoQ, which confirms that CoQ does not act as an oxido-reductant in UCP1.

Our finding that CoQ is an obligatory cofactor for H⁺ transport by UCP1 is based on reconstitution of both the native and recombinant protein, where conditions under which UCP1 acts in mitochondria may be only partially realized. However, the specific and obligatory dependence on the presence of CoQ argues for a physiological role of CoQ in the uncoupling of oxidative phosphorylation by UCP1, with several implications. First, it solves the paradox that UCP1 from inclusion bodies of *E. coli* cannot achieve its function under reconstitution conditions in which other mitochondrial carriers, such as the carriers for ketoglutarate²⁶, citrate²⁷, phosphate²⁸ and ADP/ATP (our own unpublished data) become fully active. Second, it suggests a new regulatory dimension as the uncoupling activity of UCP1 depends on the redox state of CoQ; in other words, the availability of CoQ_{ox}. There seems to be no interaction of CoQ_{red} with UCP1, as CoQ_{red} scarcely activates H⁺ transport and barely competes with CoQ_{ox}. The elucidation of such a regulation requires further studies on mitochondria. Third, with a ratio of 2.4 molecules of UCP1 to 5 molecules of CoQ per cytochrome *a*₃ complex in brown fat mitochondria¹¹, activation may also depend on a competitive recruitment of CoQ between the respiratory chain and UCP1. Fourth, unlike fatty acid, CoQ is not an H⁺ carrier and its role is expected to be more regulatory. As both are membrane bound, we presume that the molecular role of CoQ resides in cooperation with fatty acid, on the basis of a physical contact of both components at the membrane UCP1 interface. By forming a hydrogen bond between undissociated fatty acid (FAH) and the quinone oxo group, CoQ_{ox} might facilitate the entrance of FAH or of H⁺ delivered from FAH into the H⁺ channel of UCP1. Fifth, our findings are based on the paradigm role of UCP1 in the

UCP subfamily. As other UCPs are not available in native form, owing to the incomplete reconstitution of recombinant *E. coli* expressed UCPs¹⁵, widely diverging functions have been proposed^{8,9}. It is expected that the activator role of CoQ is universal to all UCPs, whether from animals or plants, as preliminary data show that CoQ is an obligatory activator of UCP2 and UCP3 expressed in *E. coli*. Last, the suggested role of UCP, particularly in plants^{7,29}, to suppress oxygen radicals generated by CoQ, and the interesting influence of CoQ derivatives on the mitochondrial permeability transition pore³⁰ now acquire a new perspective with the interaction of UCPs with CoQ. □

Methods

Expression, isolation, renaturation and reconstitution of inclusion bodies

We expressed UCP1 from hamster BAT in *E. coli* cells BL21 (DE3) transformed with the plasmid pET-24a containing the UCP1 gene. Expression was induced by isopropyl β-D-thiogalactoside. Inclusion bodies were isolated as described¹⁵. For renaturation, 5 mg of inclusion bodies were dissolved in 400 μl of 2% (w/v) sarcosyl in buffer A (50 mM NH₄HCO₃, 2 mM dithioerythritol (DTE) and 2 mM phenylmethanesulphonyl fluoride (PMSF), pH 8.0), followed by centrifugation at 13,000 g for 30 min. The supernatant was diluted by buffer A to a final sarcosyl concentration of 0.5% and protein of 3 mg ml⁻¹ and further diluted to a final concentration of 0.6 mg ml⁻¹ by buffer B (1 mM DTE, 50 μM β-mercaptoethanol, 100 mM potassium phosphate, 0.2% digitonin (water soluble form) and 0.1% sarcosyl, pH 8.0). The soluble protein was concentrated threefold by pressure dialysis at 4 °C and diluted again threefold with buffer B without sarcosyl. This step was repeated three times, and then twice with buffer C (1 mM DTE, 1 mM EDTA and 10 mM potassium phosphate, pH 8.0). The whole process took ~10 h. To remove completely sarcosyl, 1 ml of digitonin-solubilized inclusion bodies (~1.5 mg ml⁻¹) was treated with 21K Dowex mesh 20–50 (800 mg CL⁻ form and 20 mg OH⁻ form) by shaking at 4 °C overnight. The yield of renatured protein was 80–90%. The reconstitution of UCP1 in phospholipid vesicles followed essentially a described method^{15,18}.

Received 6 April; accepted 29 September 2000.

- Nicholls, D. G. & Locke, R. M. Thermogenic mechanisms in brown fat. *Physiol. Rev.* **64**, 1–64 (1984).
- Klingenberg, M. & Huang, S.-G. Structure and function of the uncoupling protein from brown adipose tissue. *Biochim. Biophys. Acta* **1415**, 271–296 (1999).
- Nedergaard, J. & Cannon, B. The uncoupling protein thermogenin and mitochondrial thermogenesis. *New Comp. Biochem.* **23**, 385–420 (1992).
- Fleury, C. *et al.* Uncoupling protein-2: a novel gene linked to obesity and hyperinsulinemia. *Nature Genet.* **15**, 269–272 (1997).
- Boss, O. *et al.* Uncoupling protein-3: a new member of the mitochondrial carrier family with tissue-specific expression. *FEBS Lett.* **408**, 39–42 (1997).
- Mao, W. *et al.* UCP4, a novel brain-specific mitochondrial protein that reduces membrane potential in mammalian cells. *FEBS Lett.* **43**, 326–330 (1999).
- Laloi, M. *et al.* Plant cold-induced uncoupling protein. *Nature* **389**, 135–136 (1997).
- Boss, O., Muzzin, P. & Giacobino, J. P. The uncoupling proteins, a review. *Eur. J. Endocrinol.* **139**, 1–9 (1998).
- Ricquier, D. & Bouillaud, F. The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP and AtUCP. *Biochem. J.* **345**, 161–179 (2000).
- Aquila, H., Link, T. A. & Klingenberg, M. Solute carriers involved in energy transfer of mitochondria form a homologous protein family. *FEBS Lett.* **212**, 1–9 (1987).
- Lin, C. S. & Klingenberg, M. Characteristics of the isolated purine nucleotide binding protein from brown fat mitochondria. *Biochemistry* **21**, 2950–2956 (1982).
- Jezek, P., Orosz, D. E., Modriansky, M. & Garlid, K. D. Transport of anions and protons by the mitochondrial uncoupling protein and its regulation by nucleotides and fatty acids. A new look at old hypotheses. *J. Biol. Chem.* **269**, 26184–26190 (1994).
- Gonzalez-Barroso, M. M., Fleury, C., Bouillaud, F., Nicholls, D. G. & Rial, E. The uncoupling protein UCP1 does not increase the proton conductance of the inner mitochondrial membrane by functioning as a fatty acid anion transporter. *J. Biol. Chem.* **273**, 15528–15532 (1998).
- Matthias, A., Jacobsson, A., Cannon, B. & Nedergaard, J. The bioenergetics of brown fat mitochondria from UCP1-ablated mice. Ucp1 is not involved in fatty acid-induced de-energization ("uncoupling"). *J. Biol. Chem.* **274**, 28150–28160 (1999).
- Echtay, K. S. *et al.* Regulation of UCP3 by nucleotides is different from regulation of UCP1. *FEBS Lett.* **450**, 8–12 (1999).

16. Murdza-Inglis, D. L. *et al.* Functional reconstitution of rat uncoupling protein following its high level expression in yeast. *J. Biol. Chem.* **266**, 11871–11875 (1991).
17. Arechaga, I. *et al.* Cysteine residues are not essential for uncoupling protein function. *Biochem. J.* **296**, 693–700 (1993).
18. Ehtay, K. S., Bienengraeber, M. & Klingenberg, M. Mutagenesis of the uncoupling protein of brown adipose tissue. Neutralization of E190 largely abolishes pH control of nucleotide binding. *Biochemistry* **36**, 8253–8260 (1997).
19. Bathgate, B., Freebairn, E. M., Greenland, A. J. & Reid, G. A. Functional expression of the rat brown adipose tissue uncoupling protein in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **6**, 363–370 (1992).
20. Jaburek, M. *et al.* Transport function and regulation of mitochondrial uncoupling proteins 2 and 3. *J. Biol. Chem.* **274**, 26003–26007 (1999).
21. Huang, S.-G. & Klingenberg, M. Fluorescent nucleotide derivatives as specific probes for the uncoupling protein: thermodynamics and kinetics of binding and the control by pH. *Biochemistry* **34**, 349–360 (1995).
22. Winkler, E., Wachter, E. & Klingenberg, M. Identification of the pH sensor for nucleotide binding in the uncoupling protein from brown adipose tissue. *Biochemistry* **36**, 148–155 (1997).
23. Skulachev, V. P. Fatty acid circuit as a physiological mechanism of uncoupling of oxidative phosphorylation. *FEBS Lett.* **294**, 158–162 (1991).
24. Brustovetsky, N. & Klingenberg, M. The reconstituted ADP/ATP carrier can mediate H⁺ transport by free fatty acids, which is further stimulated by mersalyl. *J. Biol. Chem.* **269**, 27329–27336 (1994).
25. Winkler, E. & Klingenberg, M. Effect of fatty acids on H⁺ transport activity of the reconstituted uncoupling protein. *J. Biol. Chem.* **269**, 2508–2515 (1994).
26. Fiermonte, G., Walker, J. & Palmieri, F. Abundant bacterial expression and reconstitution of an intrinsic membrane transport protein from bovine mitochondria. *Biochem. J.* **294**, 293–299 (1993).
27. Kaplan, R. S. High-level bacterial expression of mitochondrial transport proteins. *J. Bioenerg. Biomembr.* **28**, 41–47 (1996).
28. Schroers, A., Burkovski, A., Wohrlab, H. & Kramer, R. The phosphate carrier from yeast mitochondria. Dimerization is a prerequisite for function. *J. Biol. Chem.* **273**, 14269–14276 (1998).
29. Kowaltowski, A. J., Costa, A. D. & Vercesi, A. E. Activation of the potato plant uncoupling mitochondrial protein inhibits reactive oxygen species generation by the respiratory chain. *FEBS Lett.* **425**, 213–216 (1998).
30. Walter, L. *et al.* Three classes of ubiquinone analogs regulate the mitochondrial permeability transition pore through a common site. *J. Biol. Chem.* **275**, 29521–29527 (2000).

Acknowledgements

We thank K. Beyer for the NMR measurements. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to M.K.
(e-mail: klingenberg@pbm.med.uni-muenchen.de).

The PSI-H subunit of photosystem I is essential for state transitions in plant photosynthesis

Christina Lunde, Poul Erik Jensen, Anna Haldrup, Juergen Knoetzel & Henrik Vibe Scheller

Plant Biochemistry Laboratory, Department of Plant Biology,
The Royal Veterinary and Agricultural University, 40 Thorvaldsensvej,
DK-1871 Frederiksberg C, Copenhagen, Denmark

Photosynthesis in plants involves two photosystems responsible for converting light energy into redox processes. The photosystems, PSI and PSII, operate largely in series, and therefore their excitation must be balanced in order to optimize photosynthetic performance¹. When plants are exposed to illumination favouring either PSII or PSI they can redistribute excitation towards the light-limited photosystem. Long-term changes in illumination lead to changes in photosystem stoichiometry^{2,3}. In contrast, state transition is a dynamic mechanism that enables plants to respond rapidly to changes in illumination. When PSII is favoured (state 2), the redox conditions in the thylakoids change and result in activation of a protein kinase^{4–6}. The kinase phosphorylates the main light-harvesting complex (LHCII) and the mobile antenna complex is detached from PSII. It has not been clear if attachment of LHCII to PSI in state 2 is important in state transitions. Here we show that in the absence of a specific PSI subunit, PSI-H, LHCII cannot transfer energy to PSI, and state transitions are impaired.

PSI is a multi-protein complex; in higher plants, it consists of 17 different subunits⁷. The PSI reaction centre complex has a core consisting of 13 different subunits, all the electron transport factors, and about 90 molecules of chlorophyll. Surrounding the core complex are proteins of the light harvesting complex I (LHCI), which bind additional chlorophyll molecules. Four different LHCI proteins appear to be present in a fixed stoichiometry of two copies per PSI reaction centre^{7,8}. Whether LHCII normally associates with PSI in state 2 has been a controversial question. While some data have shown the PSI antenna size to increase in state 2 *in vivo*^{6,9}, there is no available information about the significance of this for state transitions. To investigate the involvement of PSI in state transitions we have generated a series of transgenic *Arabidopsis* plants lacking specific subunits of PSI. Plants lacking PSI-H, -K, -L, or -N (Δ PSI-H, Δ PSI-K, and so on) were relatively unaffected in growth, and showed nearly normal rates of photosynthesis under optimal conditions^{10–12} (Table 1). State transitions can be detected as differential changes in fluorescence from PSII when leaves are exposed alternately to light favouring PSII and light favouring PSI (Fig. 1). Plants lacking PSI-H or PSI-L were highly deficient in state transitions as evidenced in the relative fluorescence change, F_r (Fig. 1). Furthermore, non-photochemical fluorescence quenching was identical in PSII- and PSI-light. In contrast, plants lacking PSI-N or PSI-K showed almost normal state transitions. Plants with downregulation of PSI-L had a secondary loss of most of PSI-H (Table 1). As plants lacking PSI-L were slightly less affected in state transitions

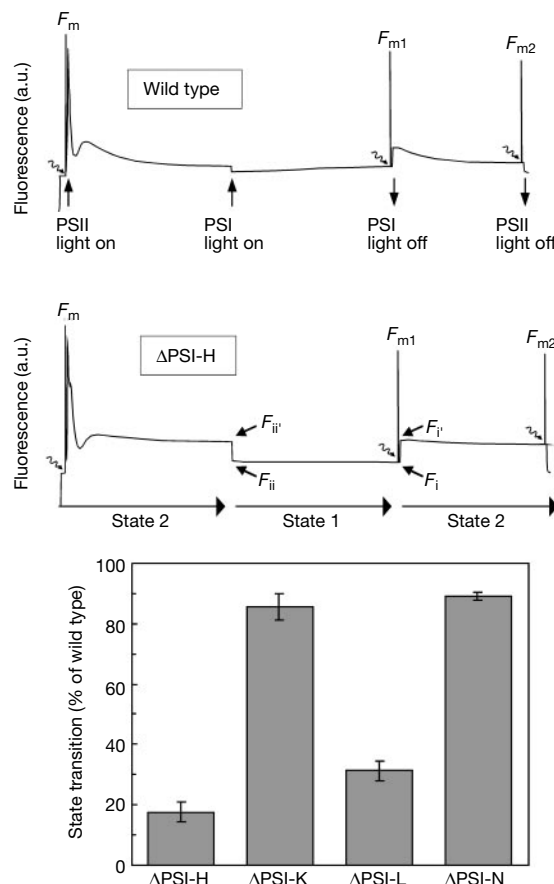


Figure 1 Measurements of transitions between state 1 and state 2. Dark-adapted leaves were exposed to either light favouring PSII (blue) or light favouring PSI (far-red). The initial high PSII fluorescence decreases as photochemical and non-photochemical quenching increase. Fluorescence decreases further when illumination includes PSI light and increases again when illumination is changed to favour PSII. Each bar represents the average ($\pm$ s.e.) for 5–6 plants. Wavy arrows point to saturating flashes (0.8 s) given 30 s before changing the PSI or PSII light.

than plants completely lacking PSI-H, we conclude that the effect on state transitions is specifically associated with PSI-H.

The lack of state transitions could have several causes: (1) docking of LHCII to PSI could be masked due to a less efficient PSI, (2) the regulatory mechanism controlling state transitions could be affected, or (3) a structural change due to the absence of PSI-H or PSI-L could cause a less efficient binding or energy transfer from LHCII to the reaction centre. NADP⁺-photoreduction measurements revealed that PSI electron transport *in vitro* is reduced by approximately 40% in the absence of PSI-H¹⁰ or PSI-L (Table 1). However, this decrease in PSI activity can not explain the lack of state transitions. In PSI lacking PSI-N, a similar reduction in electron transport was observed¹¹ but this reduction only lead to a negligible decrease in state transitions (Fig. 1).

State transitions are apparently regulated through phosphorylation of the mobile LHCII¹³. Binding of reduced plastoquinone to the Q_o site of the cytochrome *b₆f* complex activates a specific LHCII kinase in the thylakoid membrane⁶. The plastoquinone pool was more reduced in the leaves lacking PSI-H or PSI-L, as indicated by the 1-*q_p* value (Table 1; *q_p* is defined in Methods), and the kinase was functional in thylakoids prepared from plants lacking PSI-H (Fig. 2). To directly observe reversible phosphorylation *in vivo*, leaves were incubated with radioactive phosphate and irradiated with light favouring PSII or PSI (Fig. 3). Reversible phosphorylation was taking place in all plants irrespective of their ability to undergo state transitions. Therefore, the regulatory mechanisms responsible for state transitions are still intact in the absence of PSI-H or PSI-L.

It has been suggested that LHCII would donate energy to PSI through the LHCI proteins^{14–16}. Therefore, we investigated the integrity of LHCI in the transgenic plants. No change in the amounts of LHCI proteins was found in plants lacking PSI-H, -L, or -N¹¹ whereas plants lacking PSI-K¹² showed a slight reduction in LHCI proteins. Fluorescence emission spectra at 77 K from leaves lacking PSI-H did not differ significantly from spectra from wild-

type leaves, indicating that energy transfer from LHCI was essentially unaffected (data not shown). Thus, we conclude that LHCI proteins are functional and present in normal amounts in plants unable to perform state transitions. Recent structural data has indicated that the LHCI proteins bind only to the PSI-J/F side of PSI in higher plants¹⁷, that is, opposite the site where PSI-H and PSI-L are located^{18,19}. Hence, LHCI appears to dock directly to the PSI core, presumably making contact with PSI-H and possibly also PSI-L.

The structure of PSI in *Synechococcus* has been resolved at 3.5 Å resolution¹⁸. However, cyanobacteria do not contain membrane-bound LHCII and LHCI, and the PSI complex of plants contains several additional subunits. PSI-H is only found in eukaryotes and is known to be in close contact with PSI-L¹⁹. It seems likely that the addition of PSI-H in the development of eukaryotic PSI was part of the adaptation to using membrane-bound antenna complexes rather than the extrinsic phycobilisomes that are present in cyanobacteria.

By measuring the level of P700 oxidation as a function of flash intensity, the relative antenna size of PSI in state 1 and state 2 could be determined in intact leaves (Fig. 4, Table 2). Leaves lacking PSI-H showed identical PSI antenna size irrespective of the illumination, and the antenna size was identical to the wild type in state 1. In contrast, the antenna size in wild-type leaves increased 33% upon the state 1 to state 2 transition. The change in antenna in the wild type but not in leaves lacking PSI-H was confirmed by fluorescence

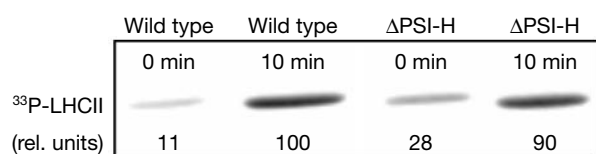


Figure 2 *In vitro* phosphorylation of LHCII. Thylakoids prepared from wild-type and ΔPSI-H plants were exposed to reducing conditions in the presence of [γ -³³P]-ATP, and the incorporation of radioactivity was followed by SDS-PAGE and phosphor-image analysis.

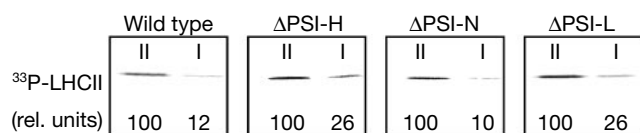


Figure 3 Reversible phosphorylation of LHCII *in vivo*. Leaves were incubated with ³³P-phosphate and exposed to PSII light (II) followed by PSI light (I). Leaf extracts were subjected to SDS-PAGE and phosphor-image analysis.

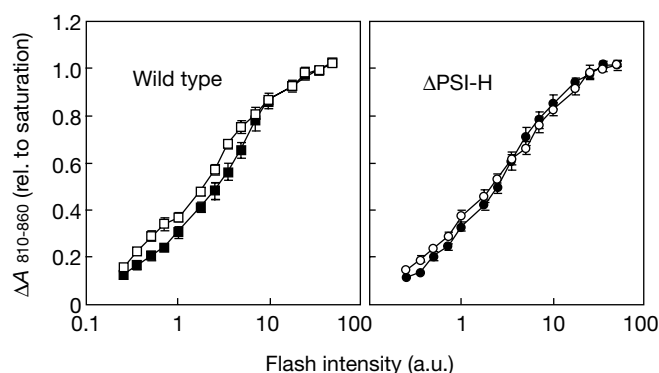


Figure 4 Light saturation curve for P700 oxidation in intact leaves. The curves were acquired by flashing the leaves with laser light of varying intensity and measuring the resulting P700 oxidation. Filled symbols represent measurements in dark-adapted plants (state 1) and open symbols represent measurements in blue light (state 2). Each curve represents an average of 12–16 leaves from four plants. Bars indicate the standard error.

Table 2 Antenna cross-section of PSI

Wild type	Wild type	PSI-H	PSI-H
State 1	State 2	State 1	State 2
100 ± 12 a	133 ± 8 b	104 ± 9 a	109 ± 10 a

The numbers are relative (wild type in state 1 = 100). Values (mean ± s.e., *n* = 12–16) followed by different letters are significantly different (*P* < 0.05).

Table 1 Photosynthetic parameters of wild-type plants and plants lacking PSI-H or PSI-L

	Wild type	ΔPSI-H	ΔPSI-L
PSI-H per P700 (% of wild type)	100	< 2	10
PSI-L per P700 (% of wild type)	100	60	< 2
<i>In vitro</i> NADP ⁺ reduction (μmol mg ⁻¹ Chl h ⁻¹)	110 ± 7 (<i>n</i> = 12)	68 ± 1 (<i>n</i> = 13)	70 ± 3 (<i>n</i> = 13)
Chl per P700	684 ± 33 (<i>n</i> = 12)	606 ± 17 (<i>n</i> = 10)	570 ± 62 (<i>n</i> = 3)
Chl <i>a/b</i>	2.55 ± 0.02 (<i>n</i> = 10)	2.66 ± 0.01 (<i>n</i> = 4)	2.66 ± 0.02 (<i>n</i> = 8)
Φ _{O₂} (μmol O ₂ μmol ⁻¹ photons)	0.112 ± 0.003 (<i>n</i> = 18)	0.102 ± 0.001 (<i>n</i> = 18)	ND
<i>P</i> (μmol O ₂ m ⁻² s ⁻¹)	9.4 ± 0.3 (<i>n</i> = 18)	8.4 ± 0.3 (<i>n</i> = 18)	ND
1 - <i>q_p</i>	0.13 ± 0.01 (<i>n</i> = 10)	0.26 ± 0.01 (<i>n</i> = 5)	0.23 ± 0.01 (<i>n</i> = 5)
<i>q_N</i> (state 2)	0.27 ± 0.02 (<i>n</i> = 10)	0.16 ± 0.01 (<i>n</i> = 5)	0.17 ± 0.01 (<i>n</i> = 5)

The amount of PSI-H and PSI-L was determined by immunoblotting. Φ_{O₂} is the apparent quantum yield of O₂ evolution. *P* is the rate of O₂ evolution in growth light. Values are means ± s.e.

excitation spectra obtained at 740 nm (data not shown). The increase in PSI antenna size in intact leaves is not significantly different from estimates based on measurements on isolated spinach thylakoids²⁰. Hence, in state 2 phosphorylated LHCII will transfer excitation energy to the PSI reaction centre in the wild type but not in plants lacking PSI-H or PSI-L. The data presented here show that not only does LHCII functionally connect to PSI in state 2, but this connection is essential for state transitions, as observed in PSII fluorescence changes. In the absence of PSI-H, non-photochemical fluorescence quenching is identical in PSI and PSII light. This suggests that LHCII in the mutant plants remains attached to PSII in spite of the unaffected LHCII phosphorylation. We therefore speculate that efficient dissociation of LHCII depends on its coupling to PSI.

As no difference in growth or appearance could be detected between the wild-type and the transgenic plants, the lack of state transitions in the latter is apparently compensated for by a higher content of PSI. Thus, under constant and optimal conditions state transitions are not essential for these plants. Two state-transition-deficient mutants have recently been identified and partly characterized in *Chlamydomonas reinhardtii*^{21,22}. Both these mutants are deficient in the phosphorylation of LHCII, and therefore the lesion is different from that reported here. The *stm1* mutant grew more poorly than wild type under most conditions, whereas the *stm7* mutant showed little effect. However, the dynamic state transitions will be more important in a natural environment with varying conditions. □

Methods

Plant material and identification of transformants

Arabidopsis thaliana Heynh. (ecotype Columbia-0) grown in Arabidopsis chambers (Percival) at 100–120 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, 8 h photoperiod, were used in all experiments. Plants lacking PSI subunits were generated by cosuppression or antisense suppression as previously described, and identified by preparing whole leaf extracts, separating proteins by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and immunoblotting with specific antibodies^{10–12}. The downregulated plants used had undetectable levels of the respective PSI subunits (< 2% of wild-type levels).

Activity measurements

NADPH⁺-photoreduction *in vitro* was determined with saturating light from absorbance changes at 340 nm (ref. 23); O₂ evolution in intact leaves was determined as described elsewhere¹¹.

State transition and antenna size measurements

Conventional fluorescence parameters were obtained as previously described¹⁰. Photochemical and non-photochemical quenching were calculated as $q_p = (F_m' - F_s)/(F_m' - F_0)$ and $q_N = (F_m - F_m')/F_m'$, respectively. F_m and F_m' designate maximal fluorescence in dark-adapted and illuminated leaves, respectively. F_s designates steady-state fluorescence under illumination, and F_0 designates fluorescence immediately after actinic illumination is turned off.

State transitions were measured on detached leaves with a pulse amplitude modulated fluorometer (Walz). The leaf was exposed to a 800-ms flash of saturating white light to determine F_m , and subsequently illuminated for 15 min with 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ blue light (PSII light) from a lamp equipped with a Corning 4-96 filter. Far-red (PSI) light (Waltz 102-FR) was turned on, and after 15 min the maximum fluorescence yield in state 1 (F_{m1}) was determined. The far-red light was then switched off and the fluorescence recorded for 15 min, after which the maximum fluorescence yield in state 2 (F_{m2}) was determined. The relative change in fluorescence was calculated as $F_r = [(F_r - F_i) - (F_{ir} - F_{i1})]/(F_i - F_0)$ (see Fig. 1), where F_i and F_{ir} designate fluorescence in the presence of PSI light in state 1 and state 2, respectively, while F_r and F_{ir} designate fluorescence in the absence of PSI light in state 1 and state 2, respectively.

The PSI antenna size in state 1 and 2 was investigated in intact leaves by determining P700 absorption change ($A_{810-860}$) with a ED-P700DW-E detector (Walz) after excitation

with a 6-ns flash from a Nd:YAG laser. Flash intensity was varied with neutral grey filters. Relative antenna size was determined by fitting the curves to a single hit Poisson distribution²⁰.

LHCII kinase activity

Thylakoids were incubated with [γ -³³P]ATP under reducing conditions²⁴. For determination of kinase activity *in vivo*, leaves were allowed to take up ³³P-phosphate for 90 min and then exposed to irradiation favouring PSII or PSI. Identical amounts of thylakoid-proteins were separated by SDS–PAGE and the degree of incorporation analysed in a phosphor-imager.

Received 17 July; accepted 29 September 2000.

1. Niyogi, K. K. Photoprotection revisited: genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 333–359 (1999).
2. Pfannschmidt, T., Nilsson, A. & Allen, J. F. Photosynthetic control of chloroplast gene expression. *Nature* **397**, 625–628 (1999).
3. Walters, R. G. & Horton, P. Acclimation of *Arabidopsis thaliana* to the light environment: changes in photosynthetic apparatus. *Planta* **195**, 248–256 (1994).
4. Allen, J. F. & Nilsson, A. Redox signalling and the structural basis of regulation of photosynthesis by protein phosphorylation. *Physiol. Plant.* **100**, 863–868 (1997).
5. Gal, A., Zer, H. & Ohad, I. Redox-controlled thylakoid protein phosphorylation. News and views. *Physiol. Plant.* **100**, 869–885 (1997).
6. Zito, F. *et al.* The Qo site of cytochrome *b₆f* complexes controls the activation of the LHCII kinase. *EMBO J.* **18**, 2961–2969 (1999).
7. Scheller, H. V., Naver, H. & Möller, B. L. Molecular aspects of photosystem I. *Physiol. Plant.* **100**, 842–851 (1997).
8. Jansson, S. A guide to the *Lhc* genes and their relatives in *Arabidopsis*. *Trends Plant Sci.* **4**, 236–240 (1999).
9. Horton, P. Are grana necessary for regulation of light harvesting? *Aust. J. Plant Physiol.* **26**, 659–669 (1999).
10. Naver, H., Haldrup, A. & Scheller, H. V. Cosuppression of photosystem I subunit PSI-H in *Arabidopsis thaliana*. *J. Biol. Chem.* **274**, 10784–10789 (1999).
11. Haldrup, A., Naver, H. & Scheller, H. V. The interaction between plastocyanin and photosystem I is inefficient in transgenic *Arabidopsis* plants lacking the PSI-N subunit of photosystem I. *Plant J.* **17**, 689–698 (1999).
12. Jensen, P. E., Gilpin, M., Knoetzel, J. & Scheller, H. V. The PSI-K subunit of photosystem I is involved in the interaction between light harvesting complex I-680 and the photosystem I reaction center core. *J. Biol. Chem.* **275**, 24701–24708 (2000).
13. Allen, J. F., Bennett, J., Steinback, K. E. & Arntzen, C. J. Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems. *Nature* **291**, 25–29 (1981).
14. Bassi, R. Spectral properties and polypeptide composition of the chlorophyll-proteins from thylakoids of granal and agranal chloroplasts of maize (*Zea mays* L.). *Carlsberg Res. Commun.* **50**, 127–143 (1985).
15. Bassi, R., Machold, O. & Simpson, D. Chlorophyll-proteins of two photosystem I preparations from maize. *Carlsberg Res. Commun.* **50**, 145–162 (1985).
16. Knoetzel, J. & Simpson, D. Expression and organisation of antenna proteins in the light- and temperature-sensitive barley mutant *chlorina*¹⁰⁴. *Planta* **185**, 111–123 (1991).
17. Boekema, E. J. *et al.* Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry* (in the press).
18. Fromme, P. Structure and function of photosystem I. *Curr. Opin. Struct. Biol.* **6**, 473–484 (1996).
19. Jansson, S., Andersen, B. & Scheller, H. V. Nearest-neighbor analysis of higher-plant photosystem I holocomplex. *Plant Physiol.* **112**, 409–420 (1996).
20. Samson, G. & Bruce, D. Complementary changes in absorption cross-sections of photosystems I and photosystems II due to phosphorylation and Mg²⁺-depletion in spinach thylakoids. *Biochim. Biophys. Acta* **1232**, 21–26 (1995).
21. Kruse, O., Nixon, P. J., Schmid, G. H. & Mullineaux, C. W. Isolation of state transition mutants of *Chlamydomonas reinhardtii* by fluorescence video imaging. *Photosynth. Res.* **61**, 43–51 (1999).
22. Fleischmann, M. M. *et al.* Isolation and characterization of photoautotrophic mutants of *Chlamydomonas reinhardtii* deficient in state transition. *J. Biol. Chem.* **274**, 30987–30994 (1999).
23. Kjær, B. & Scheller, H. V. An isolated reaction center complex from the green sulfur bacterium *Chlorobium vibrioforme* can photoreduce ferredoxin at high rates. *Photosynth. Res.* **47**, 33–39 (1996).
24. Steinback, K. E., Bose, S. & Kyle, D. J. Phosphorylation of the light-harvesting chlorophyll-protein regulates excitation energy distribution between photosystem II and photosystem I. *Arch. Biochem. Biophys.* **216**, 356–361 (1982).

Acknowledgements

We thank M. Jensen for technical assistance, B.L. Möller for discussions, and The Danish National Research Foundation for financial support.

Correspondence and requests for materials should be addressed to H.V.S. (e-mail: hvs@kvk.dk).

addendum

Magnetoresistance from quantum interference effects in ferromagnets

N. Manyala, Y. Sidis, J. F. DiTusa, G. Aeppli, D. P. Young & Z. Fisk

Nature **404**, 581–584 (2000).

Our attention has been drawn to an earlier report¹ on the zero-field resistivity of ferrimagnetic amorphous films containing mixtures of gadolinium with cobalt and iron. This paper mentions the possible relevance of electron–electron interactions to the low-temperature properties of such films, and displays the Al'tshuler–Aronov equation² for the zero-field temperature dependence of the resistivity of a disordered conductor. To understand the magnetoresistance data in our Letter, which deals with a crystalline semiconductor that undergoes a transition from insulating paramagnet to metallic magnet, a useful theoretical guidance comes from the earlier paper of Millis and Lee³, which contains explicit formulae for the field-dependent resistance. □

1. Kopelevich, Ya. V., Makarov, V. V. & Sapozhnikova, L. M. Electrical conductivity of amorphous ferrimagnetic metals. *Fiz. Tverd. Tela* **28**, 3674–3679 (1986) and *Sov. Phys. Solid State* **28**, 2069–2072 (1986).
2. Al'tshuler, B. L., Aronov, A. G., Gershenson M. E. & Sharvin, Yu. V. Quantum effects in disordered metal films. *Sov. Sci. A Phys.* **9**, 223–354 (1987).
3. Millis, A. J. & Lee, P. A. Spin-orbit and paramagnon effects on magnetoconductance and tunneling. *Phys. Rev. B* **30**, 6170–6173 (1984) and **31**, 5523–5524(E) (1985).

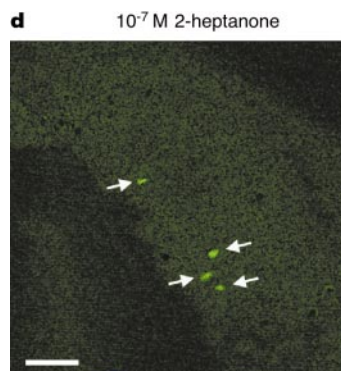
correction

Ultrasensitive pheromone detection by mammalian vomeronasal neurons

Trese Leinders-Zufall, Andrew P. Lane, Adam C. Puche, Weidong Ma, Milos V. Novotny, Michael T. Shipley & Frank Zufall

Nature **405**, 792–796 (2000).

In Fig. 3, panel f was inadvertently duplicated in panel d. The correct Fig. 3d is shown below. □



errata

Tom22 is a multifunctional organizer of the mitochondrial preprotein translocase

Sandra van Wilpe, Michael T. Ryan, Kerstin Hill, Ammy C. Maarse, Chris Meisinger, Jan Brix, Peter J. T. Dekker, Martin Moczko, Richard Wagner, Michiel Meijer, Bernard Guiard, Angelika Hönlinger & Nikolaus Pfanner

Nature **401**, 485–489 (1999).

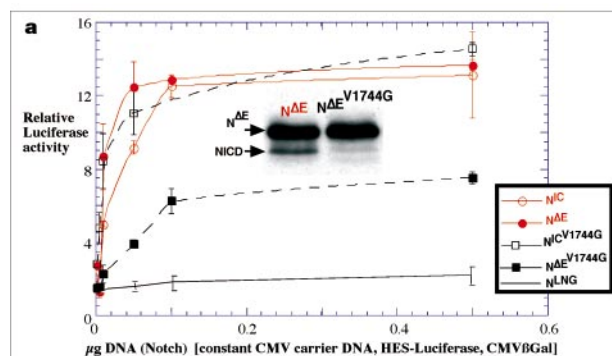
In this paper, part of the labelling of Fig. 3 was printed incorrectly. The half-tones of Fig. 3a should have been labelled from left to right: Anti-Tom22 (lanes 1 and 2); Anti-Tom40 (lanes 3 and 4); Anti-Tom20 (lanes 5 and 6); Anti-Tom70 (lanes 7 and 8). In addition, the axes of the ordinates of Fig. 3b should have been labelled from top to bottom: Co-precipitated Tom20; Co-precipitated Tom22; Co-precipitated Tom70. □

Embryonic lethality in mice homozygous for a processing-deficient allele of Notch1

Stacey S. Huppert, Anh Le, Eric H. Schroeter, Jeffrey S. Mumm, Meera T. Saxena, Laurie A. Milner & Raphael Kopan

Nature **405**, 966–970 (2000).

The contrast of Fig. 1a was too low to see all of the protein bands. Figure 1a should have appeared as below. □



Looking for new gear for 2001?

Enter a photomicrographic beauty contest, go to the library for lipids.

Poly(A) mRNA Detection

Promega www.promega.com

From picograms to nanograms

Promega's new system is designed to detect picogram to nanogram amounts of target poly(A) mRNA in RNA preparations by hybridizing anchored oligo(dT) primers to the poly(A) tails. A series of hybridization and enzymatic reactions produces a light signal, allowing the concentration of mRNA in the sample to be determined. The sensitivity limit of the system is 400 pg of mRNA, or 40 pg per μ l.

Reader Service No. 100

Reacti-Bind

Pierce www.piercenet.com

Captivating assays

Reacti-Bind anti-green fluorescent protein (anti-GFP) antibody-coated plates capture GFP fusion proteins for further assay development. The plates will bind a variety of GFP variants from crude cell extracts, allowing cell debris to be washed away. The GFP fusion protein is then properly oriented for use in other assay protocols. Reacti-bind plates are available in clear, black or white formats and are pre-blocked to save time when optimizing assay conditions.

Reader Service No. 101

MultiTEST CD45RA four-colour reagents

Becton Dickinson www.bdfacs.com

Easy as one, two, three

'Stain, lyse, no-wash' blood sample preparation is promised by the manufacturer of these reagents, which are designed for research use in immunophenotyping naive/memory T-

cell subsets and activated T-cells in flow cytometry. The reagents consist of pre-mixed combinations of four monoclonal antibodies for immunophenotyping lymphocyte subsets in whole blood. Four differently coloured fluorochromes differentiate the antibody markers. When used with the company's TruCOUNT tubes, the reagents are said to deliver absolute counts without requiring a haematology result. They are particularly suitable for investigations in the area of HIV/AIDS immunology and treatment.

Reader Service No. 102

Claudin-2 antibody

Zymed www.zymed.com

Tight junction research

Zymed announces a new polyclonal anti-claudin-2 antibody. Claudin proteins are major membrane components of tight junctional complex in epithelial and endothelial cells. Tight junctions are specialized regions of cell-cell contact that create a paracellular permeability barrier and contribute to epithelial polarity by blocking the diffusion of membrane proteins and lipids between apical and basolateral surfaces. Distinct species of claudins can interact within and between tight junctions, and thus may increase their diversity.

Reader Service No. 103

Insight Color Digital Camera

Diagnostic Instruments www.diaginc.com

New year's resolution

The Insight camera is marketed as 'the next step beyond video', producing 18 frames per second at a resolution of $1,600 \times 1,200$. Internal 10-bit sampling facilitates accurate

24-bit RGB images and microscope mounting is simplified using $1 \times C$ -mount adapters. Accompanying software enables users to view, capture, edit, annotate, measure, print and e-mail the results.

Reader Service No. 104



View finder: the Insight digital camera.

Super Hyb Kit

Molecular Research Center

www.mrcgene.com

High sensitivity in northern, southern and dot-blot hybridization

This system, which includes hybridization, pre-hybridization and wash solutions, allows detection of mRNA in RNA and eliminates the need to isolate poly A⁺ RNA. It blocks nonspecific binding of probes and augments binding to target sequences. Post-hybridization washes can be performed at room temperature. The kit can be used with both radioactive and non-radioactive probes.

Reader Service No. 105

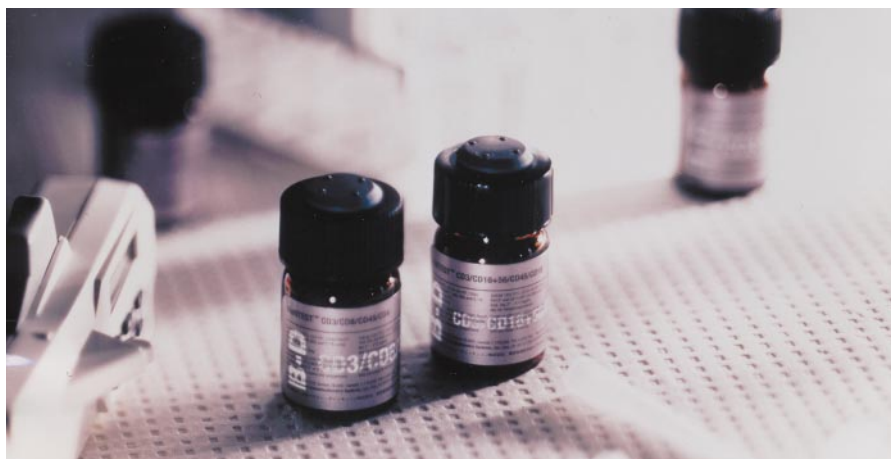
Komet Spinbar

Scienceware www.bel-art.com

One good turn deserves another

This 50-mm long stirring bar, with a diameter of 21 mm, is designed to transmit a torque load two to three times that of a conventional stirring bar. It has eight edges to generate strong turbulence. The magnet is made of samarium-cobalt, a high-energy magnetic material that reduces the risk of demagnetization, and is encapsulated in an octagonally-shaped Teflon PTFE cover. The Komet Spinbar is suitable for use in both flat and curved-bottom vessels.

Reader Service No. 106



True to type: colourful reagents for T-cell studies in flow cytometry.

new on the market

X-tremeGENE Q2

Roche www.biochem.roche.com/fugene

Expressive power

X-tremeGENE Q2 transfection reagent is optimized for transfection and high-level protein expression in HeLa, Jurkat and K-562 cells. Protein expression is claimed to be up to 10-fold higher than with other transfection reagents in Jurkat and K-562 cells without the need for stimulants such as PHA and PMA.

Reader Service No. 107

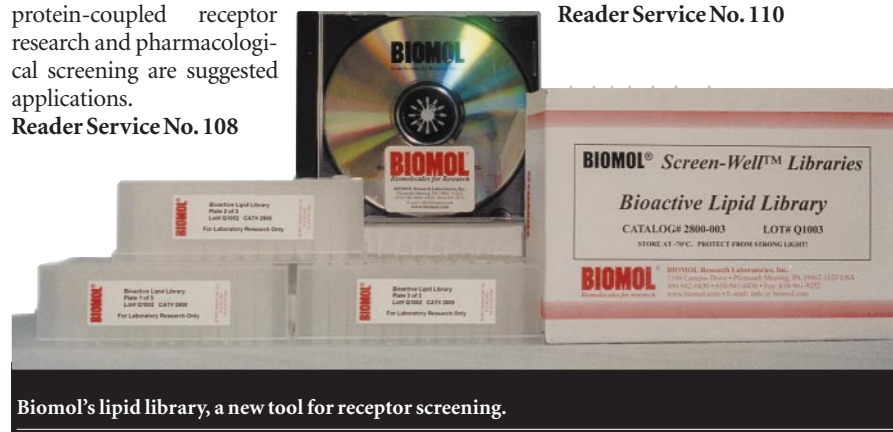
Screen-Well

Biomol www.biomol.com

Bioactive lipid library

The Screen-Well library was developed for the screening and identification of ligands for orphan receptors. It offers 200 bioactive lipids in solution at standard concentrations of 0.1 or 1.0 mM. Classes include prostaglandins, HETEs, thromboxanes, leukotrienes, LPA, lipoxins, hepoxilins, polyunsaturated fatty acids, cannabinoids, anandamides, platelet-activating factors, sphingolipids, retinoids, vitamin D metabolites, octa-decanoids, phosphatidic acids, farnesyl and geranylgeranyl derivatives; both agonists and antagonists are available. Nuclear and G protein-coupled receptor research and pharmacological screening are suggested applications.

Reader Service No. 108



Biomol's lipid library, a new tool for receptor screening.

FlaskScrubber

Labconco www.labconco.com

Residual benefits

Designed for narrow-necked laboratory glassware, the FlaskScrubber washer injects water directly to clear stubborn residue and offers both steam and multiple purified water rinses. Drying is by vacuum extraction of moisture. Temperatures can reach 60 °C, or 82 °C in the 230-volt model. Cycle duration and number of rinses can be selected using touchpad controls and progress viewed on a digital display.

Reader Service No. 109

RNAqueous

Ambion www.ambion.com

RNA isolation for RT-PCR analysis

The RNAqueous kit isolates RNA from samples as small as 1mg or 100 cells without genomic DNA contamination. The procedure yields total RNA that is treated with DNase in an optimized reaction buffer. The DNase and divalent cations are removed with a non-toxic DNase removal reagent, resulting in total RNA that is ready for RT-PCR analysis. The kit is suitable for both traditional end-point RT-PCR analysis and real-time PCR applications.

Reader Service No. 110



FlaskScrubber: end those washday blues.

Centritech

Kendro www.kendro.com

Cell separation system

Centritech gently and continuously separates live cells from dead cells and bioreactor medium utilizing a sterile, disposable insert and is claimed to be superior to perfusion systems that use hollow fibre technology. Recycling only live cells with little or no loss of viability maximizes the culture biomass, resulting in higher yields. Three models are available — two benchtop systems with throughput capacities of 0.5–10 litres per hour, and a floor-standing system for flow rates of up to 120 litres per hour.

Reader Service No. 111

Nikon Small World

Nikon Instruments www.nikonusa.com

Call for images

The Nikon International Small World Competition has been run annually by Nikon since 1975. The competition is open to any form of photograph taken through a microscope, and from this year contestants can enter by uploading digital images directly to Nikon's servers at www.nikonusa.com or www.microscopyu.com. Alternatively, an entry form can be downloaded as a pdf for submission of entries by mail. Prizes include a \$5,000 vacation or a range of Nikon products. Each entrant may submit up to three images by the deadline of 30 June 2001. Winners will be announced in November 2001, and will be featured in a Nikon calendar for 2002.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

TREVIGEN

new

PARP

Poly (ADP-ribose)
Polymerase Reagents

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com

US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

Custom Peptides & Polyclonal Antibodies

Please contact us about complete
services for your custom peptide
and polyclonal antibody needs.

Research Genetics

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 50

READER ENQUIRY NO. 36

Learning curve

Britain turns its attention to nanotechnology transfer

p620

Keeping ahead

United States backs its nanotech to the tune of \$500 million

p621

Future plans

Canada lays plans to beat the brain drain in the miniature world

p623

Opportunity knocks

Japan refuses to be left behind in the nanotech race

p624

Governments prime basic nanotech research, applied activity yet to soar

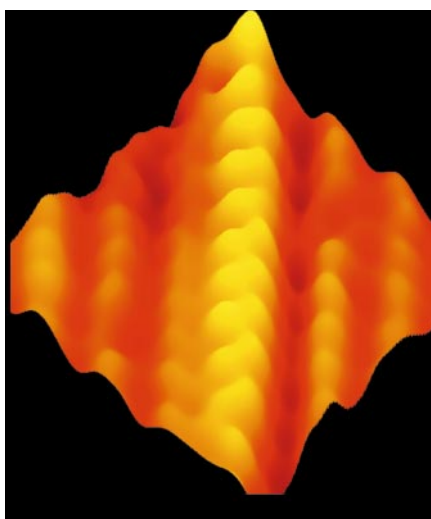
Hype or reality? Is nanotechnology just a clever rebranding, or a profound change in our world view? The answer is: both, says Helen Gavaghan.

The poet William Blake wrote of seeing "a world in a grain of sand". When Jim Gimzewski first saw the individual atoms and molecules revealed by scanning tunnelling microscopy, it truly was "like discovering a new world", he says. Gimzewski's colleagues at the IBM Research Laboratory in Zurich, Heinrich Rohrer and Gerd Binnig, invented the scanning tunnelling microscope in 1981 and won the Nobel Prize for Physics in 1986. Since then, thousands of scientists have explored the previously unseen nanocosmos — a world of atoms and molecules measuring from 0.1 to 100 nanometres.

Until the past couple of decades, it was a world that science could describe only in terms of the average behaviour and characteristics of a block of atoms and molecules. This world could be manipulated only in bulk, even though that bulk could be smaller than a grain of sand. Now, says Börje Johansson, a professor of theoretical physics at the University of Uppsala in Sweden, "we can understand at an atomic level what is happening in a chemical bond and how the molecule's function comes about". Johansson is working to develop a Swedish national strategy on nanoscience and nanotechnology.

It is this shift to understanding the nature and behaviour of matter at the scale of individual molecules and individual bonds that underpins efforts to develop nanoscale technology. This technology fascinates scientists in many disciplines. It will allow them to make medicines, materials, electronic devices and much more, in which the function is performed by an individual molecule. "Nanotechnology is the next hardware revolution in our society," says François Grey, who heads the nanotechnology group at the National Microtechnology Centre at the Technical University of Denmark in Lyngby.

In 1996, for example, Gimzewski and his colleagues succeeded in precisely repositioning one molecule. Working with collaborators at the French National Centre for Scientific Research in Toulouse and the University of Cambridge in the United Kingdom, they devised a molecule that was both cohesive



A new world: scanning tunnelling microscope images have allowed us to see the nanocosmos.

enough not to pull apart when pushed by the tip of a scanning tunnelling microscope and able to slide over a surface without sticking to the microscope's tip. Although other researchers had achieved this in environments cooled to -270°C to reduce thermal vibrations, Gimzewski and his colleagues worked at room temperature, an important step towards constructing a single molecule with a specific function.

So something real is happening. Yet scientists all over the world have for years been working at the nanoscale level in materials science, physics, chemistry or biology. In industry, developments in biotechnology

About 80% of what is being talked about as nanotechnology is marketing, and 20% is real.

and semiconductors approach the nanoscale level. And then, says Grey, "one day all these people woke up and discovered that what they were really doing was nanotechnology".

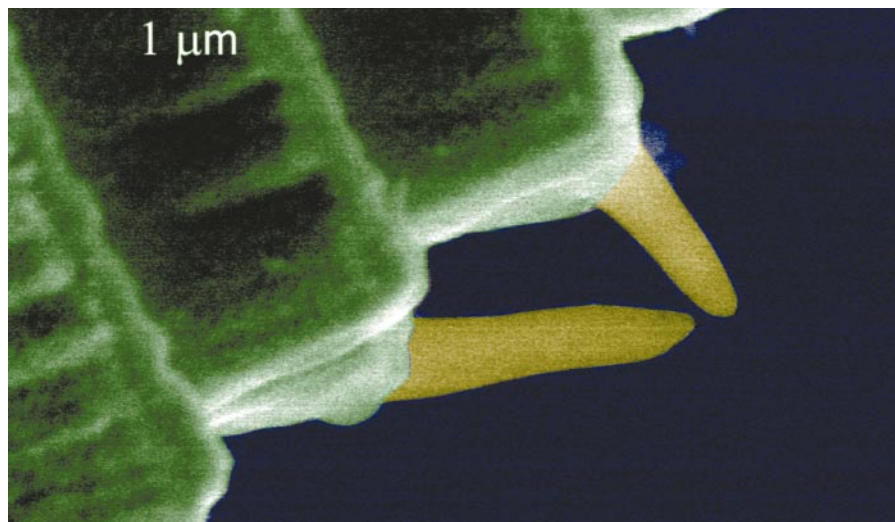
It is a catchy word, and not everyone who has adopted it is actually developing technologies that assign function to individual molecules rather than the traditional bulk properties. Nevertheless, the word helps scientists to attract funders and entrepreneurs to win financing for start-up companies.

"Nanotechnology is a sexy name," says Jean-Christophe Eloy from the international marketing and consultancy group, Yole Développement, which advises high-technology companies. "I'd say 80% of what is being talked about as nanotechnology is marketing, and 20% is real."

Some fields are closer to turning nanoscience into nanotechnology than others, notably the semiconductor industry and the biotechnology sector. But researchers are also developing dot-sized lasers for optical communications, constructing nanowires and catalysts that promote reactions at the level of a single molecule, or designing supramolecules, with sites for more than one function. For example, one subset of molecules might act as a sensor to detect a protein sequence on a cancer cell, then once the supramolecule is attached, another site could initiate a toxic reaction to kill the cell.

What these examples have in common is a strong interdisciplinary nature. Recognizing that nanotechnology breaks the boundaries of traditional disciplines is the key to finding a job in this newly named field. "I'd say that a physicist needs to make sure they know some chemistry, or a chemist knows some physics," says Peter Bøggild from the Danish National Microtechnology Centre.

Bøggild is currently recruiting for a project to develop nanoscale 'tweezers' that can move an individual molecule. He and collaborating teams in Japan, the United States, Sweden and Denmark will explore how to use these in materials science, surface science and biology. "Candidates need to be willing to learn," says Bøggild, "and they need an



Get a grip: nanotweezers could allow researchers to manipulate individual atoms routinely.

open mind scientifically and socially.”

They also need to be willing to move around the world if they want a job, because the field is more developed in some places than others. They will need to stay alert, because the job may not be labelled nanotechnology. Many academic projects are still funded within classical disciplines and have not rebranded themselves. But increasingly, that is changing, not only as the word gains popularity but as it becomes shorthand for a genuine shift in understanding.

Within the European Union (EU) there is a range of plans for nanotechnology. The most advanced is the Nanotechnology Information Devices (NID) initiative, which funds collaborative research aimed at developing new processing and memory systems at the atomic or molecular scale. Unlike most EU projects to date, the NID funds long-term speculative research and does not insist on large numbers of collaborators.

Perhaps the most ambitious EU plan is that of research commissioner Philippe Busquin, to establish a European Research Area coordinating member states' basic scientific research efforts, with nanotechnology as a priority. But at present, the truest picture of Europe's efforts in nanotechnology is spending by national governments.

Some EU states have been more focused than others. France has a network of university and research centres focused on nanotechnology. Germany established six nanotechnology competence centres in 1998: hubs of virtual networks linking academic and industrial groups in each of six broad areas of activity. One network comprises 28 companies, 10 universities or institutes and 15 research centres, all with research efforts in ultrathin materials. Potential applications include sensors, artificial skin, electronic components and corrosion protection.

Each hub is staffed by two or three people, to promote networking, further the public understanding of nanotechnology, advise on

education and provide a link with the German research ministry. It is a measure of Germany's confidence in the area that not all the country's leading research organizations have even felt the need to join these centres of competence.

Anyone with a good proposal, whether affiliated to a centre of competence or not, can apply for a portion of the DM60 million (\$US26 million) that the German government is setting aside annually for nanotechnology research. Gerd Bachmann, one of the main architects of the government's strategy for research and development in nanoscience and nanotechnology, says the country probably spends as much again on nanotechnology projects that go by some other name.

In Germany, says Bachmann, it is not lack of money or intent that might hold up nanotechnology in the long run, but a shortage of good-quality students studying physics and chemistry. As in most other parts of western

Europe, students opt for less demanding topics. “It can be a nightmare to find the skilled people,” says Eloy, “but they can be the key to a company's development.” Despite this shortage of students, scientists interviewed by *Nature* say there is no shortage of good applicants for academic and industrial posts in nanotechnology.

Important work is also being done outside the EU. Switzerland became the first European country to focus on nanoscience when its National Science Foundation launched a research programme in 1986 on the physics and chemistry of surfaces. Swiss efforts are now on the brink of changing from science to technology. To encourage the trend, the Swiss Federal Institute of Technology has set up a five-year, SFr60 million (\$US33 million) programme to fund nanotechnology with industrial applications. The funds must be matched 50:50 by industry.

“From the beginning we wanted to involve industry, but that proved difficult,” says Hans Hintermann, a professor of materials science and project director of the National Science Foundation's national research programmes in Berne. “Now I have the feeling, not only in Switzerland, but when I listen to colleagues in other countries, that the time is right for industry.”

Helen Gavaghan is a science and technology writer based in Hebden Bridge, West Yorkshire.

Nanotechnology Information Devices initiative

♦ <http://www.cordis.lu/ist/fetnid.htm#intro>

Technology Roadmap for Nanoelectronics

♦ <http://www.cordis.lu/esprit/src/melna-rm.htm>

European NanoForum

♦ <http://www.nanoForum.de>

Nanotechnology competence centres

♦ <http://www.nanonet.de/english/default.htm>

The Swiss nanotech site

♦ <http://www.nanoscience.ch>

French nanotech initiatives

♦ <http://www.rmnt.org>

Britain turns its attention to nanotech transfer

The United Kingdom realized some time ago that it is lagging behind in the race to develop and commercialize nanotechnology. Low public awareness and confusion about the potential of the technology have meant that progress has been slow compared with many other countries.

“Each year we see an increasing number of researchers from diverse disciplines entering the field of nanotechnology and, although new companies are being spun out, it is not at the rate seen in the United States and other European countries,” says Ottilia

Saxl, director of Britain's Institute of Nanotechnology (IoN).

But the government has announced its intention to increase its nanotech funding, back more interdisciplinary research, and work harder to commercialize academic nanotech research — all recommendations made in the recent IoN report *Opportunities for Industry in the Application of Nanotechnology*.

Established in 1997, the IoN acts as a bridge between academic researchers and companies. In the past, the UK government did not actively encourage academic

A flexible approach



Andrew Turberfield may not be a typical scientist, but the University of Oxford physicist's journey from academia to industry and then back again — with detours in biology and engineering — may resemble a common roadmap for nanotechnologists.

Turberfield, a semiconductor physicist by training, decided to spend 1998 on sabbatical at Lucent Technologies' Bell Labs in New Jersey because he was "keen to learn some biophysics". When he actually got to work, he decided to look at the research from a different perspective — using a biological system for engineering purposes.

His approach was very successful; he talks fondly of the year he spent there working on molecular machines made from DNA. He describes the machines as tweezers with stiff arms

seven nanometres long made from double-stranded DNA. Selectively adding complementary strands of DNA can open and close those arms.

Back in Oxford he is building a molecular motor that will move along a track. He is also investigating general strategies for nanostructure fabrication based on molecular recognition of DNA strands. The DNA motors could be used to control self-assembly of microscopic components, for example to create molecular-scale integrated circuits.

Turberfield is now developing an interdisciplinary research programme in collaboration with colleagues in the physical and life sciences departments. His alliance with Lucent is also still strong and his work with them continues.

A. J.

► <http://www-nt.physics.ox.ac.uk/qop>

researchers to commercialize their work, says Saxl, who co-authored the IoN report. But, she adds, Britain's commercial world needs nanotech. The IoN points out that Britain's pharmaceutical sector (which accounts for 6% of world sales and a third of Britain's research and development) and its rapidly growing biotechnology industry are obvious benefactors of nanotechnology research. They would find immediate use for the potential technologies such as nanoparticles in drug delivery.

Technology transfer in nanotechnology can strengthen the United Kingdom's overall economy — especially by funding technologies that can be imported. According to an estimate by the Royal Academy of Engineering, around 60% of all diagnostic and monitoring instrumentation used by the National Health Service originates in other countries.

Saxl believes that Britain can learn to accelerate its technology transfer by looking abroad. The IoN is arranging for a team of key academics from Britain to visit the United States and Germany early next year. Both these countries have already been spending a considerable amount on nanotechnology research. The academics aim to bring back information on nanotechnology policy and how nanotechnology is being successfully applied.

The United Kingdom can also strengthen its nanotech efforts by involving scientists from many different fields, says Adam Curtis, at the Centre for Cell Engineering at Glasgow University. "There is a big opportunity for interdisciplinary research — electrical engineers, chemists and biologists will all make up future nanotechnology researchers," he says. The recent joint call for proposals from several of the United Kingdom's research councils indicates that the potential funders are already thinking in this direction. Curtis is part of a consortium bidding for funds to establish two interdisciplinary research centres for nanotechnology.

But George Smith, head of the materials department at the University of Oxford, calls

the UK nanotech funding "hopelessly inadequate" compared with funding in other countries. The government's plans for two interdisciplinary centres funded with around £10 million (US\$14 million) over five years each pale in comparison with the US National Nanotechnology Initiative, which has a budget of \$500 million for 2001 alone. The comparison gives "some idea of

the scale of the mountain we have to climb", Smith says.

Alok Jha is a reporter in London for *Research Fortnight*.

Institute of Nanotechnology

► <http://www.nano.org.uk/>

UK Research Councils

► <http://www.research-councils.ac.uk/>

Centre for Cell Engineering, Glasgow University

► <http://www.gla.ac.uk/centres/cellengineering>

US industry starts to think big by acting small

Nanotechnology now resembles biotechnology or information technology in their infancies — before venture capital flowed unimpeded, and jobs abounded. But predicting job growth in this developing field is difficult, not least because nanotechnology is more than one thing.

The broadest definition describes nanotechnology as the manipulation of individual molecules smaller than 100 nanometres (although some would argue for 50 nanometres). The concept has sparked scientific imaginations because working at this 'nanoscale' changes the physical properties of materials — making things really small provides researchers with another way to control materials.

That prospect of greater control opens up options in many areas. In materials science, for example, a polymer could be made more durable by mixing it with nanoparticles. Particles fused to pieces of DNA could be made into medical devices to diagnose different diseases. And in nanoelectronics, tiny circuits could be built up molecule by molecule, rather than be etched away using standard lithography techniques.

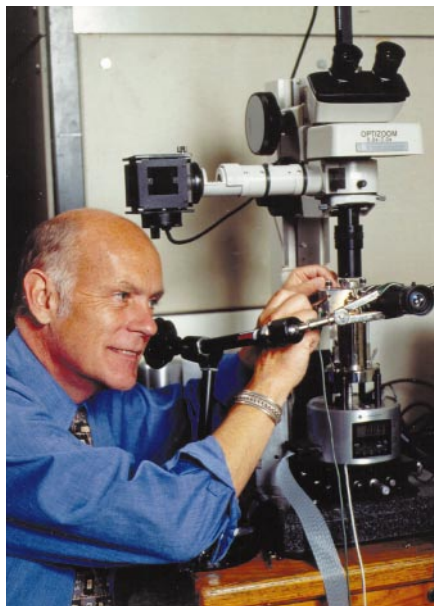
Job prospects in each of these areas will differ. But to jumpstart the opportunities, the US National Science Foundation (NSF) has doubled its budget for nanotech-

nology research, to about \$500 million in 2001.

But this rise must be kept in perspective, says Stanley Williams, director of the Quantum Structures Research Initiative at Hewlett-Packard Laboratories in Palo Alto, California, who wrote a report on nanotechnology with Mihail Roco, programme director of the NSF's nanotechnology initiative. He believes the \$250 million increase to be "a drop in the bucket", especially when it is compared with the entire budgets of other US science funding agencies, such as the National Institutes of Health, which has seen recent hikes in excess of \$1 billion. He feels that although the initiative will boost the currently low success rate for people applying for NSF grants in nanotechnology, it will still only be able to fund about a fifth of the proposals it receives.

And new jobs will not emerge immediately, Williams adds. "Are there a lot of jobs and do we see a lot of jobs in the next year or so? I would say no," he says. He suspects that nanotechnology will first transform existing jobs. But if the technology proves successful, nanotech jobs could follow the same exponential growth as was seen when information technology and biotechnology took off.

Williams believes materials science is the area likely to benefit first. The field is already



Tubular sells: Richard Smalley hopes to capitalize on nanotubes.

▶ following the biotech model: innovations in universities, followed by spin-off companies, and eventually adaptation by larger firms.

Richard Smalley at Rice University is a case in point. Smalley shared the 1996 Nobel Prize for Chemistry for discovering buckminsterfullerene (C_{60}) — a new form of carbon — and its attendant family of fullerenes. Since then, he has begun manufacturing a tubular version of the carbon macromolecules. He is now set to step up production with his new company, Carbon Nanotechnologies.

Initially, the company will remain small, at around 25 employees. It will focus on producing tubes for researchers who might use them to make polymers, medical devices or to build circuits. The company might eventually try some of those applications on its

own. “But right now we just want to make the stuff,” says Smalley.

Dow Chemical, on the other hand, is testing the potential of using clay nanoparticles in polymers. The company has a small programme — \$16 million over five years, half of which comes from the US National Institute of Standards and Technology — to design plastics that evenly incorporate nanoparticles.

“The target here is to design plastics that have the behaviour of metals,” says Juan Garces, a senior research scientist at Dow. He hopes to be able to coax the particles to self-assemble and so disperse evenly into the polymer matrix. Because the clay nanoparticles have a higher ratio of atoms on their surface than do larger clay molecules, he expects the composite formed to withstand higher temperatures than conventional polymers. But expanding this programme so that it generates more jobs relies on more than just scientific success — it must also make money.

Instrumental technology

One application for nanotechnology that is potentially profitable is in medical testing devices, notably those that can read small amounts of DNA. This has attracted university researchers and resulted in a number of spin-offs.

Chad Mirkin, director of Northwestern University’s Institute for Nanotechnology incorporated his company, Nanosphere (based in Evanston, Illinois), last January, out of his research on using DNA to aid self-assembly. The company, which has eight employees and is on track to double this figure in a year, aims to produce particles that will bind specifically to minute quantities of DNA, which could then be used as very sensitive probes for disease.

One of the company’s staff scientists, James Storhoff, says the company epitomizes

how truly multidisciplinary nanotechnology can be. Physical chemists look at the properties of different materials to work out which will make the best nanoparticle for a particular application. Organic chemists then find ways to attach the DNA pieces to the particles. Biologists look primarily at the DNA to



James Storhoff recognizes the interdisciplinary nature of nanotech.

find out what sequences would be best for diagnosing particular conditions. And engineers will eventually build a device that scans the particle and identifies the presence of disease.

One of the most talked about areas of nanotechnology holds arguably the greatest promise of long-term job prospects: developing computer chips with ever-smaller circuitry. “Lithography is going to reach a limit,” says Thomas Theis, head of physical sciences at IBM’s Thomas J. Watson Center in Yorktown, New York.

So instead of carving away at a chip, the next step will be to build circuitry from the ground up, perhaps using DNA to put the switches in place. IBM has probably the largest nanocomputing groups in the world with 40–60 ‘hardcore’ nanotechnology researchers.

IBM built such a large team because the company realized that processing speed and storage space could not keep increasing without a technical shift. The same impetus is driving DuPont iTechnologies, which makes material for integrated circuits, and Agilent, to invest more in basic research.

Agilent is rebuilding its basic research group after splitting with Hewlett-Packard last November. The group functions more like an academic team, and is judged more on publication than on product, says Len Cutler, director of precision instrumentation and basic research at Agilent in Palo Alto. The team now has seven members, but the company would like to double that. Agilent, like so many other companies, is poised to take advantage of the nanotech boom, but remains cautious of investing too much too soon.

Paul Smaglik is Nature’s correspondent in Washington DC and Careers and Recruitment editor.

National Nanotechnology Initiative

♦ <http://www.nano.gov>

Carbon Nanotechnologies

♦ <http://www.cnanotech.com>

Nanosphere

♦ <http://www.nanosphere-inc.com>

Nanogen

♦ <http://www.nanogen.com/www1/index.htm>

Building for the future

If universities are to be the engines for powering nanotechnology, then new facilities may be the sparkplug. So far only one institution, Rice University, has what Stanley Williams, a senior principal lab scientist at Hewlett-Packard Laboratories in California, calls true nanotechnology capabilities — the resources to produce and analyse materials smaller than 100 nanometres.

But several universities are planning to build such facilities. Harvard’s proposal drew Charles Marcus from a tenured position at Stanford, even though the Harvard project has not yet broken ground. Although the National Science Foundation funds several nanofabrication centres, “you have to wait in line”, says Marcus.

Northwestern University’s Institute for Nanotechnology will get its own \$34 million building in 2002. And the University of California at Los Angeles, partnering with the University of California at Santa Barbara, is hoping to receive state funds for a \$360 million facility. Williams says the completion of those buildings will coincide with when he expects the job market in nanotechnology to take off — the next five to ten years. Just as biotech companies spun out of the boom of life science research over the past ten years, he expects nanotech companies to emerge from university nanotech programmes as they continue to expand.

P.S.

Center For Nanoscale Science And Technology at Rice University

♦ <http://cnst.rice.edu>

Northwestern University Institute for Nanotechnology

♦ <http://www.northwestern.edu/research/nanotechnology/index.html>

A compendium of US nanofabrication facilities

♦ http://deas.harvard.edu/~jones/lab_arch/nano_facilities/nano_facilities.html

Canada tries to limit nanotech brain drain

Canada has traditionally lost some of its best and brightest minds to the United States, mainly because it cannot compete with the resources of an economy ten times the size of its own. And nanotechnology has been no exception.

Martin Moskovits recently moved from the chair of the University of Toronto's chemistry department to become dean in mathematics, life and physical sciences at the University of California at Santa Barbara. "In nanotechnology in Canada, we're slow off the mark," he says. "We haven't quite analysed appropriately what's made hot technological economies like the United States and Europe great."

More funding for basic research, with less constraints on practical outcomes, has stoked the field in the United States. And Moskovits hopes that Canada will adopt such an approach to staunch the flow of researchers south.

Scientists at Canada's National Research Council (NRC) are working to do just that. They are pushing the government to create a nanotechnology institute, and are hoping that the recent federal research funding largesse might allow them to succeed. The Liberal government increased funds for research and innovation in its February budget by Can\$4.1 billion (US\$2.6 billion). And in its election campaign material, it has promised to "at least double" those expenditures (see *Nature* 404, 8; 2000).

Robert Wolkow, leader of the NRC's Scanning Tunneling Microscopy Group, says a large part of Canada's talent in the nanoworld is at the NRC, but these researchers will not benefit from the recent funding bonanza because the increases are directed at strengthening universities. "It's tragic," Wolkow says. "It's already leading to the departure of some really hot people."

With the support of others, Wolkow is



Despite working abroad, Martin Moskovits hopes that Canada can make its mark in nanotech.

"pushing for a radically different kind of NRC institute, where globally competitive salaries would be raised by some imaginative means and the very best [scientists attracted], not just from Canada but from everywhere."

Wolkow, a Canadian who, after graduating from the University of Toronto joined IBM in New York for a postdoc, had the good fortune to work with one of the first scanning tunnelling microscopes. He did the first study that showed a chemical reaction on an atom-by-atom scale, which got him a job at Lucent Technologies' Bell Labs in New Jersey. There he built the first tuneable-temperature cryogenic scanning tunnelling microscope, which uses surface temperature to freeze reactions or allow them to continue at a controlled rate.

Despite the lack of a strong infrastructure, Canadian scientists have continued to advance nanotechnology. Wolkow and colleagues recently created one-dimensional organic structures on a silicon surface by

combining scanning tunnelling microscopy and self-directed growth techniques. They hope that this technique could pave the way for assembling molecular wires on the nanotech equivalent of microchips.

There are other bright spots on the Canadian horizon. Nobel laureate John Polanyi and postdoctoral associates Ping-He Lu and Duncan Rogers at the University of Toronto have developed a technique for imprinting the structures of molecules on underlying substrates such as silicon. Work which could lead to methods for making microscopic structures for molecular electronics.

In eastern Canada, Bruce Balcom at the University of New Brunswick has adapted magnetic resonance imaging (MRI) to observe and quantify changes taking place in materials in space and over time. He says the university's MRI Research Centre is the only one in Canada and the only university-based laboratory of its kind in North America. Its techniques can be used with biological structures *in vivo* and with a large range of materials such as concrete, polymers, composites and microporous solids.

Harry Ruda, who directs the Energenius Centre for Advanced Nanotechnology (ECAN) at the University of Toronto, is a leader in so-called collective effects leading to quantum computing systems. ECAN was founded in 1997 with matching Can\$45 million awards from private sources, the Ontario government and the university. The centre carries out research and training in semiconductor nanotechnology, and has joint projects with the NRC, Cornell University and the University of North Carolina at Chapel Hill.

Ruda says that the centre is different because of its interdisciplinary focus on fabrication methods for nanostructures, and in exploring and exploiting collective effects to make novel electronic and photonic devices. ECAN seeks applications from graduate students and post-doctoral researchers, and encourages its members to participate in international exchange programmes.

Although Moskovits is now in the United States, he maintains a Canadian connection. He leads a programme in molecular electronics for the Canadian Institute for Advanced Research, involving self-assembly and synthesis of nanoelectronic materials. The institute is a 'university without walls', and serves to reduce the brain drain by enlisting the talents of researchers worldwide — including Canadians who have left Canada. ■

David Spurgeon is *Nature's* Canada correspondent.

MRI Research Centre

♦ www.utoronto.ca/physics/mri/index.html

Energenius Centre for Advanced Nanotechnology

♦ www.utoronto.ca/~ecan

NRC Steacie Institute for Molecular Sciences

♦ www.nrc.ca/sims

NRC Institute for Molecular Sciences

♦ www.nrc.ca/ims

Nanotech undergrad course available

The University of Toronto will next year offer a nanoengineering undergraduate programme that combines chemistry, physics and materials science engineering.

Douglas Perovic, professor of applied science and engineering, says that the programme has already sparked a huge response. Students will be recruited from the university's engineering science course, but others can apply. The curriculum will open up four new professorships, as well as new positions in electrical computer engineering, physics and chemistry.

The new course involves seven university units: the materials science, physics, chemistry, chemical engineering, applied chemistry, mechanical and industrial engineering departments, and the institute for biomaterials and biomedical engineering. During the first two years, students will concentrate heavily on maths and physics, but in the final two years they will specialize in nanoengineering-type subjects. In the fourth year they will prepare a research thesis, as is done in masters programmes. In addition, they will be able to select groups of courses in their individual fields of interest.

D. S.

Japan sets sights on success in nanotechnology

As the United States launches its drive to capitalize on nanotechnology's potential, Japan is also gearing up for expansion in the field. Extra money has been earmarked for a host of new multidisciplinary programmes, centres and institutes for nanoscience. All of which is good news for researchers looking for work in the country.

But with expansion comes the possibility of organizational disarray. Some researchers describe Japan's current nanoscience programmes as chaotic. So the multitude of new programmes and institutes threatens to add to the madness as well as to the opportunity. By the end of the year, a high-level committee under the Council for Science and Technology will set out proposals covering how the various programmes should be coordinated.

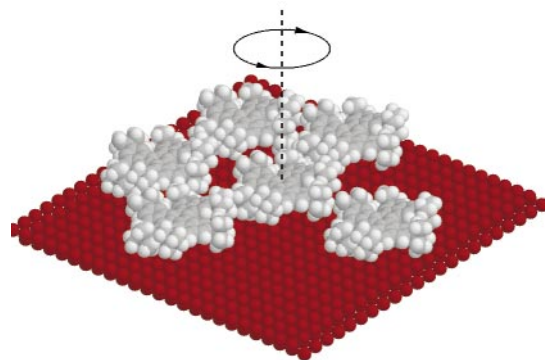
It could be a mammoth task. The reorganized National Institute of Advanced Industrial Science and Technology (AIST), for example, will have six centres and four institutes spread around the country, all taking a

multidisciplinary approach to nanoscience and nanotechnology. The Correlated Electron Research Center will explore new electronic materials for developing atom-based technology. For example, it will address questions of how to make quantum nanostructures and device junctions using oxide materials. "The physics of these exploratory electronic materials is not yet established," says director Yoshinori Tokura.

About 40 scientists will staff the centre, half of whom will be postdocs. Masataka Hirose of Hiroshima University will run AIST's Advanced Semiconductor Research Center, with significant cooperation from the private sector and universities. The centre will address fundamental problems in making semiconductors smaller than is possible using existing chip technology. It aims to reduce integrated circuit technology to 50–70 nanometres by 2007.

Plans for the future

The National Institute for Materials Sciences (NIMS), to be formed from a merger of two Science and Technology Agency institutes, will undertake a range of research on materials that could be used for nanodevices and nanoscale environmental energy projects. Three likely areas of research, according to NIMS director Kazuhiro Yoshihara, are laser generation using quantum dots on semiconductors, single-electron transistors, and computers using



In a spin: nanotechnology could revolutionize many areas of science.

quantum dots or quantum wires. The institute will be housed in a new building and will have a yearly budget of ¥2.4 billion (US\$22 million). Yoshihara expects the centre to employ around 100 people, of which about 60 will be taken from the predecessor institutions.

The New Energy and Industrial Technology Development Organization (NEDO) is supporting a separate nanomaterial programme. This is receiving yearly funds of ¥5 billion for 5–7 years, which programme co-leader Teruo Kishi hopes will be increased after next year. The programme will be split into eight projects, primarily representing different materials such as polymers and particles. Six project leaders will come from universities and two from the AIST. The project's 150–200 permanent staff will be joined by 100–200 postdocs and temporary employees. Despite the fact that foreigners will probably not be candidates for the permanent positions, Kishi hopes to attract many postdocs from abroad.

University-led nanotech programmes appear poised to flourish as well. For example, Tadahiro Ohmi at Tohoku University aims to reduce quantum noise and statistical fluctuation as integrated circuits approach 50 nanometres and then move on towards 10 nanometres. The 50-nanometre level is already heavily supported with funds and researchers from industry, but Ohmi is looking for partners to move beyond that. Although it is hard to get funding for postdocs at a university, he is hoping to obtain greater industry support and so increase the number of postdocs on his project from two to ten within two or three years. There are also plans for a research institute focusing on future silicon technologies.

David Cyranoski is *Nature's* Japan correspondent.

Agency of Industrial Science and Technology (precursor to National Institute of Advanced Industrial Science and Technology)

♦ <http://www.aist.go.jp>

New Energy and Industrial Technology Development Organization

♦ <http://www.nedo.go.jp>

National Research Institute for Metals (precursor to National Institute of Materials Sciences)

♦ <http://www.nrim.go.jp:8080/public/english>

If you have a clear idea of what you want to do in your research, you can get almost anything.

Land of opportunity

Permanent positions for foreigners are still exceedingly rare in Japan. But British-born, Oxford-trained Suzi Jarvis is one who has turned a string of grants into a six-year stay. Starting next year, she will become one of the few foreign researchers in the country to run their own lab. She will be based in Tsukuba at what is provisionally called the Nanotechnology Laboratory.

One of Japan's key attractions for foreign researchers is its tremendously well-equipped laboratories. Jarvis is working on nanoscale interactions and has an ultra-high vacuum and a liquid atomic-force microscope more or less to herself — a luxury almost unknown in other labs around the world. "If you have a clear idea of what you want to do in your research, you can get almost anything you want in terms of materials and instruments," she says.

"I've never had such freedom and resources," agrees Jose A. Torres, who researches the structural and conduction properties of gold nanowires, and has worked at several European research institutes. Indeed, Japanese researchers who travel abroad are shocked at the sparseness of some of the top US university labs.

The abundance and generosity of research grants and positions in Japan are also a draw. Salaries for postdoc researchers at the Agency of Industrial Science and Technology usually range between ¥6 million and ¥8 million (US\$54,000–72,500). And according to Jarvis, postdoctoral grants from the New Energy and Industrial Technology Development Organization pay about US\$65,000 a year — outdoing some professor's salaries in Britain — with a good deal of holiday and conference travel included.

Yet these grants are undersubscribed, and many directors of Japan's new institutes are already complaining of difficulties in getting well-trained international researchers to come to Japan. Part of the reason has to do with lack of awareness of opportunities in Japan, but there are also some cultural walls that stand in the way.

D.C.